(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 October 2002 (03.10.2002)

PCT

(10) International Publication Number
WO 02/077157 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/11257

(22) International Filing Date: 22 March 2002 (22.03.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/278,336 23 March 2001 (23.03.2001) US

(71) Applicant *(for all designated States except US)*: **SYNGENTA PARTICIPATIONS AG** [CII/CII]; Schwarzwaldallee 215, CH-4058 Basel (CH).

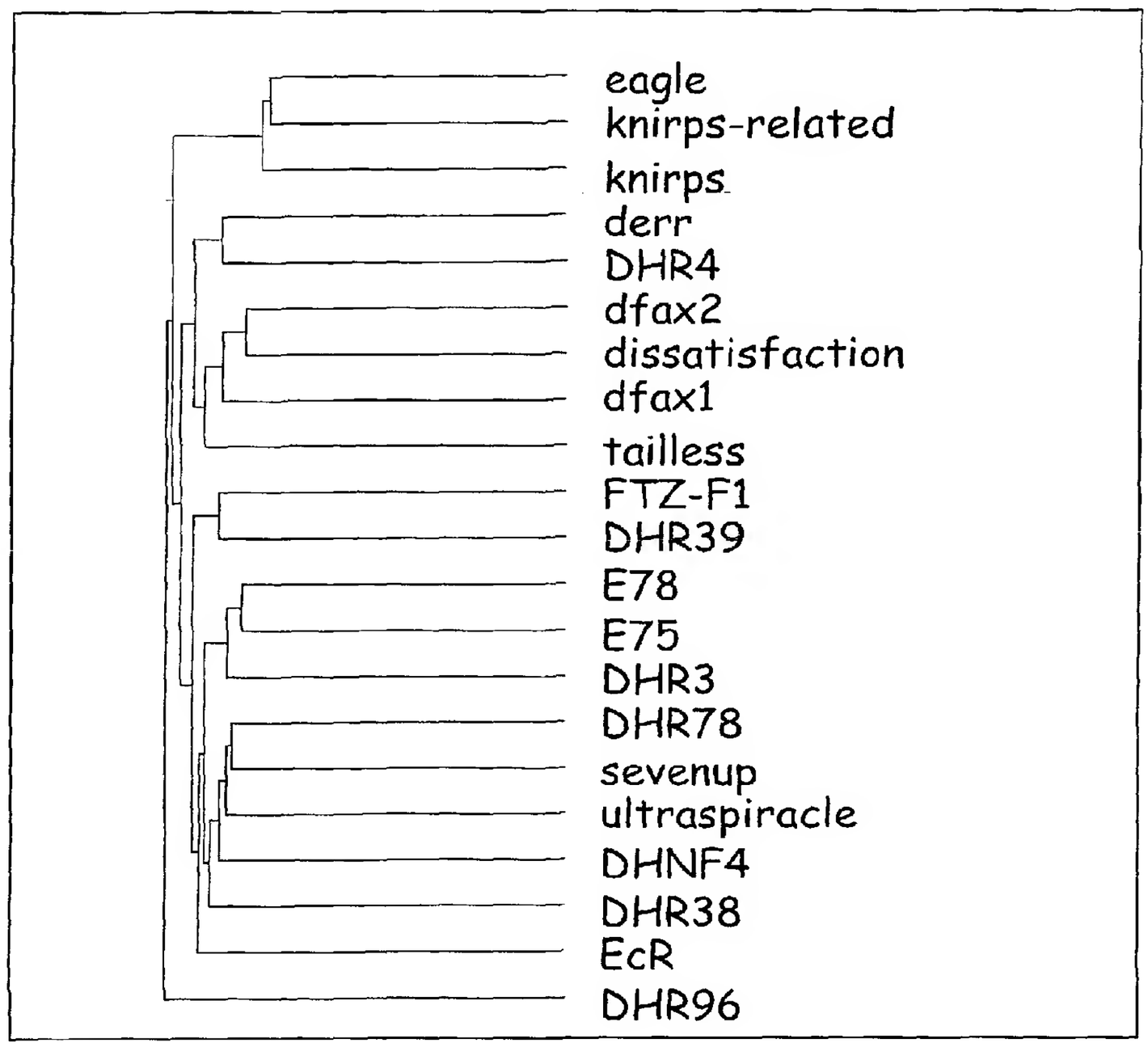
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BROADUS, Julie, A.** [US/US]; 13304 Drew Hill Road, Chapel Hill, NC 27514 (US). **BROWN, Blanche, Covington** [US/US]; 209 Creeks Edge, Chapel Hill, NC 27516 (US). **STAM, Lynn, F.** [US/US]; 2409 Pleasant Union Church Road, Raleigh, NC 27614 (US). **KAMDAR, Kim, P.** [US/US]; 906 Muirlands Vista Way, La Jolla, CA 92037 (US).

(74) Agent: **TAYLOR, Arles, A., Jr.**; Jenkins & Wilson, P.A., University Tower, Suite 1400, 3100 Tower Boulevard, Durham, NC 27707 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

[Continued on next page]

(54) Title: INSECT NUCLEAR RECEPTOR GENES AND USES THEREOF



(57) Abstract: The present invention provides isolated nucleic acids encoding insect nuclear receptor polypeptides, isolated insect nuclear polypeptides, and uses thereof. The disclosed insect nuclear receptor nucleic acids and polypepides can be used in screening assays to identify insecticidal compounds. The disclosed insect nuclear receptor nucleic acids and polypeptides can further be used as components of a chimeric expression cassette for inducible gene expression.

WO 02/077157 A2

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Description

INSECT NUCLEAR RECEPTOR GENES AND USES THEREOF

Field of the Invention

The present invention generally relates to nuclear receptor genes. More particularly, the present invention provides novel nuclear receptor nucleic acid and polypeptide sequences, chimeric genes comprising the disclosed nuclear receptor sequences, antibodies that specifically recognize the disclosed nuclear receptor polypeptides, modulators of nuclear receptor nucleic acids and polypeptides, and uses thereof.

Table of Abbreviations

ATCC	American Tissue Culture Collection
βFTZ-F1	fushi tarazu transcription factor 1, β isoform
CDS	coding sequence
DBD	DNA-binding domain
DERR	*Drosophila* estrogen-related receptor
DFAX1	*Drosophila fax*-related gene 1
DFAX2	*Drosophila fax*-related gene 2
DHR38	*Drosophila* hormone receptor 38
DHR39	*Drosophila* hormone receptor 39
DHR4	*Drosophila* hormone receptor 4
DSF	dissatisfaction
dsRNA	double-stranded RNA
dsRNAi	double-stranded RNA interference
E75A	ecdysone-inducible gene E75, A isoform
EcR	ecdysone receptor
EGON	eagle nuclear receptor
DHR3	*Drosophila* hormone receptor 3
DHR78	*Drosophila* hormone receptor 78
DHR96	*Drosophila* hormone receptor 96

E78	*Drosophila* ecdysone-inducible protein 78
EGON	eagle
FCS	Fluorescence Correlation Spectroscopy
GST	glutathione S transferase
HMM	Hidden Markov Model
DHNF4	*Drosophila* hepatic nuclear factor 4
HR3	hormone receptor 3
hs	*Drosophila hsp70* promoter
IPM	integrated pest management
JH	juvenile hormone
LBD	ligand-binding domain
PCR	polymerase chain reaction
PEG	polyethylene glycol
PTTH	prothoracicotropic hormone
RACE	rapid amplification of cDNA ends
SELDI-TOF MS	Surface-Enhanced Laser Desorption/ Ionization Time-of-Flight Mass Spectroscopy
Sf9 cells	*Spodoptera frugiperda* cells
SPR	Surface Plasmon Resonance
USP	ultraspiracle

Background Art

Insects contribute or cause many human and animal diseases, and are responsible for substantial agricultural and property damage. The societal costs associated with insect pests in dollars, time, and suffering are monumental. To combat these problems, insecticidal compounds have been developed and employed. The total worldwide market size for insecticide crop protection is over $5 billion, and insecticide products comprise approximately 32% of world consumption of pesticides.

Insecticide development has been guided predominantly by leadfinding efforts for new chemical structures. According to this strategy, chemical

derivatization of a known insecticide is performed, and the synthesized compounds are analyzed for insecticidal activity. An alternative approach relies on methods for detecting molecular interactions between a candidate compound and a target molecule. An ideal target molecule is precisely regulated during insect development, such that modulation of the activity or level of activity of the target molecule results in organismal lethality. High throughput screening methods have enabled rapid screening of diverse and populous compound libraries for an ability to interact with a target molecule. The novel modulators discovered by such methods are useful as insecticides.

A target molecule can be further selected based on modulation of the target molecule activity that results in lethality during larval development. The insect life cycle requires successive larval or nymph stages that are devoted to growth such that the animal can increase mass by several thousand–fold. To sustain this growth, immature insects feed unabated for prolonged periods, and thus are particularly deleterious to agricultural crops during this developmental stage.

Each larval instar concludes with molting of the larval cuticle to accommodate the changing size of the larvae. The apolysis of the old cuticle and the synthesis of new cuticle are regulated by the coordinate action of juvenile hormone and ecdysone (20-hydroxyecdysone, hereafter referred to as ecdysone). The neuropeptide PTTH directs a transient rise in hemolymph titer of ecdysone, which is the trigger for molting. A concomitant high level of juvenile hormone signals resynthesis of larval cuticle, while low juvenile hormone levels signal the synthesis of pupal cuticle and commitment to metamorphosis. *See* Nijhout (1994) Insect Hormones, Princeton University Press, Princeton, New Jersey.

Classical endocrinology studies demonstrated that a premature rise in hemolymph ecdysone achieved by feeding larvae ecdysone was sufficient to trigger premature molting. More recently, experiments in *Drosophila* have provided genetic evidence for ecdysone control of larval molting. Mutation of the *ultraspiracle* (*usp*) gene, which encodes the heterodimeric partner of EcR

to form the functional ecdysone receptor, results in larval lethality with supernumerary spiracles indicative of an incomplete molt (Perrimon et al. (1985) *Genetics* 11:23-41; Oro et al. (1992) *Development* 115(2):449-462). The *ecdysoneless* gene encodes a protein required for ecdysone synthesis, and animals carrying mutations in *ecdysoneless* show defective larval molting and death (Henrich et al. (1987) *Dev Biol* 120(1):50-55; Henrich et al. (1993) *Dev Genet* 14(5):469-477). Similarly, mutations of the *dare* and *dre4* genes, which encode proteins required for ecdysone synthesis, result in defective molting (Freeman et al. (1999) *Development* 126:4591-4602; Sliter & Gilbert (1992) *Genetics* 130:555-568). These studies demonstrate that misregulated ecdysone signaling – either premature or absent – disrupts larval molting and leads to larval death.

Molecular cloning of the ecdysone receptor has provided a foundation for understanding the mechanism of ecdysone signaling and the mode of action for growth regulatory insecticides. The functional ecdysone receptor is a heterodimer of the ecdysone receptor (EcR) and ultraspiracle (USP) nuclear receptor proteins (Koelle et al. (1991) *Cell* 67:59-77; Koelle (1992) Ph.D. Thesis, Stanford University, Stanford, California; Yao (1993) *Cell* 71:63-72; Thomas et al. (1993) *Nature* 362:471-475). Ecdysone enters cells by virtue of its hydrophobic structure, and binds the EcR/USP heterodimer. The hormone-bound or activated receptor binds DNA at regulatory sequences called response elements and therein directs gene transcription. A subset of the immediate gene targets of the ecdysone signal, the primary response genes, are transcription factors that turn on a large group of secondary response genes that ultimately lead to altered cell functions. Thus, the hormone signal elicits a transcription cascade that both amplifies and diversifies the initial signal.

Several insecticides are known to elicit insect lethality by interfering with nuclear receptor signaling. The non-steroidal ecdysone agonists RH5849 (Wing (1988) *Science* 241:467-469), RH2485 (methoxyfenozide, Dhadialla et al. (1998) *Annu Rev Entom* 43: 545-569), and RH5992 (tebufenozide, also known as the insecticide MIMIC®), are chemical ligands

for the ecdysone receptor (EcR). *See also*, Dhadialla et al. (1998) *Annu Rev Entom* 43: 545-569, incorporated herein by reference, which describes several insecticides with ecdysteroidal and juvenile hormone activity.

At least seven other nuclear receptor genes are regulated by ecdysone, implicating their participation in larval growth and molting as well. Thummel (1997) *BioEssays* 19(8):669-672. At a molecular level, these nuclear receptors might function in yet undiscovered hormone signaling pathways. Alternatively or in addition, they might modulate ecdysone signaling by heterodimerization and transregulation. *See* White et al. (1997) *Science* 276(5309:114-117; Sutherland et al. (1995) *Proc Natl Acad Sci USA* 92(17):7966-7970.

Functional analysis of orphan nuclear receptors can be addressed using standard molecular and genetic techniques in *Drosophila melanogaster* (herein after "*Drosophila*"). In addition to EcR and USP, fourteen nuclear receptor genes have been identified in *Drosophila*: *knirps*, *knirps-related* (*knrl*), *egon/eagle (eg)*, *seven-up* (*svp*), *tailless* (*tll*), *hepatic nuclear factor 4* (*DHNF4*), *fushi tarazu factor 1* (*FTZ-F1*), *ecdysone-inducible protein 75* (*E75*), *ecdysone-inducible protein 78* (*E78*), *hormone receptor 3* (*DHR3*), *hormone receptor 38* (*DHR38*), *hormone receptor 39* (*DHR39*), *hormone receptor 78* (*DHR78*), and *hormone receptor 96* (*DHR96*) (reviewed in Thummel (1995) *Cell* 83:1-20). All such receptors, with the exception of EcR and USP, are designated orphan nuclear receptors to signify that corresponding endogenous ligands have not been identified. The insect proteins KNIRPS, KNIRPS-RELATED, and EGON lack a discernible ligand binding domain, but are classified as nuclear receptors based on the distinctive sequence of their DNA binding domains (Nauber et al. (1988) *Nature* 336(6198):489-492; Oro et al. (1988) *Nature* 336(61989):493-496; Higashijima et al. (1996) *Development* 122(2):527-536).

There exists a continuing demand for insecticides that show improved efficacy and new modes of action. To this end, the present invention discloses a functional characterization of nuclear receptors during

Drosophila larval development. Nuclear receptors that confer larval lethality when misregulated are used in biochemical assays as targets for insecticide development. The present invention also discloses several novel insect nuclear receptor polypeptides and nucleic acid molecules encoding the same that are further useful as components of gene switch technology for inducible gene expression.

Summary of the Invention

The present invention discloses isolated insect nuclear receptor polypeptides and isolated nucleic acid molecules encoding the same. Preferably, an isolated insect nuclear receptor polypeptide, or functional portion thereof, comprises a polypeptide encoded by the nucleic acid molecule of any one of SEQ ID NOs:1, 5, 9, 13, 17, 19, 21, 23, and 25; a polypeptide encoded by a nucleic acid molecule that is substantially identical to any one of SEQ ID NOs:1, 5, 9, 13, 17, 19, 21, 23, and 25; a polypeptide having an amino acid sequence of any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26; a polypeptide that is a biological equivalent of any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26; or a polypeptide that is immunologically cross-reactive with an antibody that shows specific binding with a polypeptide comprising some or all amino acids of any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26.

The present invention further teaches chimeric genes having a heterologous promoter that drives expression of a nucleic acid sequence encoding an insect nuclear receptor polypeptide. Preferably, the chimeric gene is carried in a vector and introduced into a host cell so that an insect nuclear receptor polypeptide of the present invention is produced. Preferred host cells include but are not limited to a bacterial cell, an insect cell, and a plant cell.

In another aspect of the invention, a method is provided for detecting a nucleic acid molecule that encodes an insect nuclear receptor polypeptide. According to the method, a biological sample having nucleic acid material is hybridized under stringent hybridization conditions to an insect nuclear

receptor nucleic acid molecule of the present invention. Such hybridization enables a nucleic acid molecule of the biological sample and the insect nuclear receptor nucleic acid molecule to form a detectable duplex structure. Preferably, the insect nuclear receptor nucleic acid molecule includes some or all nucleotides of any one of SEQ ID NOs:1, 5, 9, 13, 17, 19, 21, 23, and 25. The present invention further teaches an antibody that specifically recognizes an insect nuclear receptor polypeptide. Preferably, the antibody recognizes some or all amino acids of any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26. A method for producing an insect nuclear receptor antibody is also disclosed, and the method comprises recombinantly or synthetically producing an insect nuclear receptor polypeptide, or portion thereof, as set forth in any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26; formulating the insect nuclear receptor polypeptide so that it is an effective immunogen; immunizing an animal with the formulated polypeptide to generate an immune response that includes production of insect nuclear receptor antibodies; and collecting blood serum from the immunized animal containing antibodies that specifically recognize an insect nuclear receptor polypeptide. Antibody-producing cells can be optionally fused with an immortal cell line whereby a monoclonal antibody that specifically recognizes an insect nuclear receptor polypeptide can be selected.

A method is also provided for detecting a level of insect nuclear receptor polypeptide using an antibody that recognizes an insect nuclear receptor polypeptide of any of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26. According to the method, a biological sample is obtained from an experimental subject and a control subject, and an insect nuclear receptor polypeptide is detected in the sample by immunochemical reaction with the insect nuclear receptor antibody. Preferably, the antibody recognizes amino acids of any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26; and is prepared according to a method of the present invention for producing such an antibody.

The present invention further discloses a method for identifying a compound that modulates nuclear receptor function. The method

comprises: (a) exposing an isolated insect nuclear receptor polypeptide of any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, 26 to one or more compounds, and (b) assaying binding of a compound to the isolated insect nuclear receptor polypeptide. A compound is selected that demonstrates specific binding to the isolated insect nuclear receptor polypeptide. Preferably, the modulator is a chemical compound, a protein, a peptide, a nucleic acid, or an antibody, and was prepared according to a method disclosed herein.

The present invention also provides a method for identifying an insecticidal compound that modulates nuclear receptor function. The method comprises: (a) isolating an insect nuclear receptor polypeptide of any one of even numbered SEQ ID NOs:2-34, wherein modulation of the insect nuclear receptor polypeptide confers lethality of an insect during a larval stage; (b) exposing the isolated insect nuclear receptor polypeptide to a plurality of substances; (c) assaying binding of a substance to the isolated nuclear receptor polypeptide; and (d) selecting a substance that demonstrates specific binding to the isolated insect nuclear receptor polypeptide. Preferably, the modulator is a chemical compound, a protein, a peptide, a nucleic acid, or an antibody, and was prepared according to a method disclosed herein.

The present invention further provides a method for preventing or treating an insect infestation of a plant, the method comprising: (a) preparing an insecticidal composition that is a modulator of an insect nuclear receptor set forth as any one of even-numbered SEQ ID NOs:2-34; and (b) contacting an effective dose of the insecticidal composition with a plant, whereby an insect infestation of the plant is prevented or abrogated. Preferably, the insecticidal composition comprises a chemical compound, a protein, a peptide, a nucleic acid, or an antibody, and was prepared according to a method disclosed herein. Preferably, the insect infestation is abrogated by lethality of the insect. In one embodiment, the insecticidal composition also displays nematicide activity, such that contacting an effective dose of the

insecticidal composition with a plant prevents or abrogates a nematode infestation of the plant.

The present invention further provides a method for preventing or abrogating an insect infestation of a plant, the method comprising: (a) expressing in a plant an insect nuclear receptor modulator that modulates the activity of an insect nuclear receptor polypeptide of any one of even-numbered SEQ ID NOs:2-34, whereby an insect infestation of a plant is prevented or abrogated. Preferably, the insecticidal composition comprises a protein, a peptide, a nucleic acid, or an antibody. In one embodiment, the insecticidal composition additionally displays nematicidal activity, such that expression of insect nuclear receptor modulator in a plant prevents or abrogates a nematode infestation of the plant. The present invention further embodies plants, plant tissues, plant seeds, and plant cells that express an insect nuclear receptor modulator and that are therefore able to inhibit plant parasitic infestation.

The present invention also discloses a chimeric nuclear receptor cassette comprising a DNA binding domain, a ligand binding domain, a hinge domain, and an activation or repression domain, wherein one or more of the DNA binding domain, ligand binding domain, hinge domain, or activation of repression domain is identical or substantially identical to a portion of any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26.

Also disclosed is a method of inducing expression of a target nucleic acid sequence. In a preferred embodiment, the method comprises: (a) constructing a chimeric nuclear receptor expression cassette wherein one or more of the DNA binding domain, ligand binding domain, and activation/repression domains is identical or substantially identical to a portion of any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26; (b) constructing a target expression cassette having a target nucleotide sequence and a cis-regulatory element that is recognized by a DNA binding domain of the chimeric nuclear; (c) expressing the chimeric nuclear receptor expression cassette and the target expression cassette in a heterologous organism; and (d) contacting a ligand that binds to the ligand binding domain

of the chimeric nuclear receptor with the organism, whereby the target nucleotide sequence is expressed. In one embodiment, the method is performed to induce gene expression in a plant. The present invention also encompasses plants, plant tissues, plant seeds, and plant cells comprising a disclosed nuclear receptor expression cassette and a target expression cassette.

Accordingly, it is an object of the present invention to provide novel insect nuclear receptor nucleic acids and polypeptides, and novel methods relating thereto. This object is achieved in whole or in part by the present invention.

An object of the invention having been stated above, other objects and advantages of the present invention will become apparent to those skilled in the art after a study of the following description of the invention, Figures, and non-limiting Examples.

Brief Description of the Drawings

Figure 1 is a neighbor-joining tree generated using the pileup feature of the GCG sequence analysis program (Devereux et al. (1984) *Nuc Acids Res* 12:387-395). The tree depicts relationships among the homeodomain regions of *Drosophila* nuclear receptors (eagle, GenBank Accession No. D43634; knirps-related, GenBank Accession No. X14153; knirps, GenBank Accession No. X13331; DHR4, GenBank Accession No. AL035245 and SEQ ID NO:14; dissatisfaction, GenBank Accession No. AF106677; tailless, GenBank Accession No. AF019362; FTZ-F1, GenBank Accession No. M98397; DHR39, GenBank Accession No. L07551; E78, GenBank Accession No. U01087; E75, GenBank Accession No. X51548; DHR3, GenBank Accession No. M90806; DHR78, GenBank Accession No. U36791; sevenup, GenBank Accession No. M28863; USP, GenBank Accession No. X53417; DHNF4, GenBank Accession No. U70874; DHR38, GenBank Accession No. X89246; EcR, GenBank Accession No. M74078; DHR96, GenBank Accession No. U36792; DERR, SEQ ID NO:2; DFAX1, SEQ ID NO:6; DFAX2, SEQ ID NO:10).

Figures 2A-2B presents an alignment among USP proteins derived from the indicated insect species. B.mori, *Bombyx mori* USP (GenBank Accession No. AAC13750; SEQ ID NO:43); M.sexta, *Manduca sexta* USP (GenBank Accession No. AAB64234; SEQ ID NO:44); C.fumiferana, *Choristoneura fumiferana* USP (GenBank Accession No. AAC31795; SEQ ID NO:45); H.virescens, *Heliothis virescens* USP (SEQ ID NO:20); L.migratoria, *Locusta migratoria* USP (GenBank Accession No. AAF00981; SEQ ID NO:46); C.tentans, *Chironomus tentans* USP (GenBank Accession No. AAC03056; SEQ ID NO:47); D.melanogaster, *Drosophila melanogaster* USP (GenBank Accession No. X53417; SEQ ID NO:48). The core DBD is underlined.

Figures 3A-3B presents an alignment among FTZ-F1 proteins derived from the indicated insect species. hv.bftz, *Heliothis* FTZ-F1 (SEQ ID NO:18); bmftz, *Bombyx mori* FTZ-F1 (GenBank Accession No. P49867; SEQ ID NO:49); dmftz, *Drosophila* βFTZ-F1 (GenBank Accession No. M98397; SEQ ID NO:32).

Figures 4A-4F presents an alignment among E75 proteins derived from the indicated insect species. D.melanogaster (A), *Drosophila* E75A (GenBank Accession No. A34598; SEQ ID NO:34); M.sexta (A), *Manduca sexta* E75A (GenBank Accession No. Q08893; SEQ ID NO:50); D.melanogaster (B), *Drosophila* E75B (GenBank Accession No. B34598; SEQ ID NO:51); M.sexta (B), *Manduca sexta* E75B (GenBank Accession No. C56591; SEQ ID NO:52); D.melanogaster (C), *Drosophila* E75C (GenBank Accession No. P13055; SEQ ID NO:53); C. fumiferana, *Choristoneura fumiferana* E75 (GenBank Accession No. O01639; SEQ ID NO:54); G. mellonella, *Galleria mellonella* E75 (GenBank Accession No. P50239; SEQ ID NO:55); M.ensis, *Metapenaeus ensis* E75 (GenBank Accession No. AAC71770; SEQ ID NO:56); H.virescens, *Heliothis virescens* (SEQ ID NO:18). The core DBD is underlined.

Figure 5 is a bar graph that depicts percentage survival of *Drosophila* following injection of dsRNA corresponding to the indicated nuclear receptors as described in Example 5 herein below. Solid bar, DHR3; gray

bar, DHR4; open bar, EGON; cross-hatched bar, FTZ-F1; wavy bar, DSF; checkerboard bar, DERR; stippled bar, FAX1; vertical line bar, FAX2; horizontal line bar, buffer (control).

Figure 6 is a bar graph that depicts *Drosophila* larval lethality induced by overexpression of the indicated nuclear receptors. Gray bars, larvae that were not heat treated; solid bars, larvae that were heat treated at 0-2 hours post-hatching; open bars, larvae that were heat treated at 20-22 hours post-hatching; control, w^{1118} larvae; DHR38, *yw*; P[*hs-DHR38* w^{+}]-II larvae; DHR39, w^{1118}; P[*hs-DHR39-6* w^{+}] or w^{1118}; P[*hs-DHR39-3* w^{+}] larvae; E75A, w^{1118}; +/SM5; P[*hs-E75A* w^{+}]/TM3 or w^{1118}; +/SM5; P[*hs-E75A* w^{+}]/TM3 larvae. Error bars indicate standard deviation.

Brief Description of Sequences in the Sequence Listing

Odd-numbered SEQ ID NOs:1-31 are nucleotide sequences described in Table 1.

Even-numbered SEQ ID NOs:2-32 are protein sequences encoded by the immediately preceding nucleotide sequence, *e.g.*, SEQ ID NO:2 is the protein encoded by the nucleotide sequence of SEQ ID NO:1, SEQ ID NO:4 is the protein encoded by the nucleotide sequence of SEQ ID NO:3, etc.

SEQ ID NOs:35-42 are PCR primers.

SEQ ID NOs:43-56 are protein sequences available from GenBank that are presented in the Figures for comparison with novel sequences disclosed herein.

Table 1. Sequence Listing Summary

SEQ ID NO.	description
1-2	*Drosophila* gm estroten-related receptor (gmDERR)
3-4	*Drosophila* estrogen-related receptor (DERR)
5-6	*Drosophila* fax-related receptor 1 (gmDFAX1)
7-8	*Drosophila* fax-related receptor 1 (DFAX1)
9-10	*Drosophila* gm fax-related receptor 2 (gmDFAX2)
11-12	*Drosophila* fax-related receptor 2 (DFAX2)
13-14	*Drosophila* hormone receptor 4 (DHR4)
15-16	*Drosophila* dissatisfaction (DSF)
17-18	*Heliothis* fushi tarazu factor 1 (FTZ-F1)
19-20	*Heliothis* ecdysone-inducible protein 75 (E75)
21-22	*Heliothis* ultraspiracle (USP)
23-24	*Heliothis* hepatic nuclear factor 4 (HNF4)
25-26	*Heliothis* hormone receptor 3 (HR3)
27-28	*Drosophila* hormone receptor 38 (DHR38)
29-30	*Drosophila* hormone receptor 39 (DHR39)
31-32	*Drosophila* fushi tarazu factor 1, β isoform (βFTZ-F1)
33-34	*Drosophila E75A*
35	βFTZ-F1 degenerate forward primer A
36	βFTZ-F1 degenerate forward primer B
37	βFTZ-F1 degenerate reverse primer A
38	βFTZ-F1 degenerate reverse primer B
39	βFTZ-F1 forward primer A
40	βFTZ-F1 forward primer B
41	βFTZ-F1 reverse primer A
42	βFTZ-F1 reverse primer B
43	*Bombyx* ultraspiracle (USP)
44	*Manduca* ultraspiracle (USP)
45	*Chorisoneura* ultraspiracle (USP)
46	*Locusta* ultraspiracle (USP)
47	*Chironomus* ultraspiracle (USP)
48	*Drosophila* ultraspiracle (USP)
49	*Bombyx* FTZ-F1
50	*Manduca* E75A
51	*Drosophila* E75B
52	*Manduca* E75B
53	*Drosophila* E75C
54	*Choristoneura* E75
55	*Galleria* E75
56	*Matapenaeus* E75

Detailed Description of the Invention

I. Definitions

While the following terms are believed to be well understood by one of ordinary skill in the art, the following definitions are set forth to facilitate explanation of the invention.

I.A. Nucleic acids

The nucleic acid molecules provided by the present invention include the isolated nucleic acid molecules of any one of SEQ ID NOs:1, 5, 9, 13, 17, 19, 21, 23, and 25; sequences substantially identical to sequences of any one of SEQ ID NOs:1, 5, 9, 13, 17, 19, 21, 23, and 25; conservative variants thereof, subsequences and elongated sequences thereof, complementary DNA molecules, and corresponding RNA molecules. The present invention also encompasses genes, cDNAs, chimeric genes, and vectors comprising disclosed nuclear receptor nucleic acid sequences.

The term "nucleic acid molecule" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses nucleic acids containing known analogues of natural nucleotides that have similar properties as the reference natural nucleic acid. Unless otherwise indicated, a particular nucleotide sequence also implicitly encompasses conservatively modified variants thereof *(e.g.*, degenerate codon substitutions), complementary sequences, subsequences, elongated sequences, as well as the sequence explicitly indicated. The terms “nucleic acid molecule” or “nucleotide sequence” can also be used in place of “gene”, “cDNA”, or “mRNA”. Nucleic acids can be derived from any source, including any organism.

The term “isolated”, as used in the context of a nucleic acid molecule, indicates that the nucleic acid molecule exists apart from its native environment and is not a product of nature. An isolated DNA molecule can exist in a purified form or can exist in a non-native environment such as a transgenic host cell.

The term “purified”, when applied to a nucleic acid, denotes that the nucleic acid is essentially free of other cellular components with which it is associated in the natural state. Preferably, a purified nucleic acid molecule is a homogeneous dry or aqueous solution. The term "purified" denotes that a nucleic acid gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid is at least about 50% pure, more preferably at least about 85% pure, and most preferably at least about 99% pure.

The term "substantially identical”, in the context of two nucleotide sequences, refers to two or more sequences or subsequences that have at least 60%, preferably about 70%, more preferably about 80%, more preferably about 90-95%, and most preferably about 99% nucleotide identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms (described herein below under the heading “Nucleotide and Amino Acid Sequence Comparisons” or by visual inspection. Preferably, the substantial identity exists in nucleotide sequences of at least 50 residues, more preferably in nucleotide sequence of at least about 100 residues, more preferably in nucleotide sequences of at least about 150 residues, and most preferably in nucleotide sequences comprising complete coding sequences. In one aspect, polymorphic sequences can be substantially identical sequences. The term “polymorphic” refers to the occurrence of two or more genetically determined alternative sequences or alleles in a population. An allelic difference can be as small as one base pair.

Another indication that two nucleotide sequences are substantially identical is that the two molecules specifically or substantially hybridize to each other under stringent conditions. In the context of nucleic acid hybridization, two nucleic acid sequences being compared can be designated a “probe” and a “target”. A “probe” is a reference nucleic acid molecule, and a “‘target” is a test nucleic acid molecule, often found within a heterogeneous population of nucleic acid molecules. A “target sequence” is synonymous with a “test sequence”.

A preferred nucleotide sequence employed for hybridization studies or assays includes probe sequences that are complementary to or mimic at least an about 14 to 40 nucleotide sequence of a nucleic acid molecule of the present invention. Preferably, probes comprise 14 to 20 nucleotides, or even longer where desired, such as 30, 40, 50, 60, 100, 200, 300, or 500 nucleotides or up to the full length of any of those set forth as SEQ ID NOs:1, 5, 9, 13, 17, 19, 21, 23, and 25. Such fragments can be readily prepared by, for example, directly synthesizing the fragment by chemical synthesis, by application of nucleic acid amplification technology, or by introducing selected sequences into recombinant vectors for recombinant production.

The phrase "hybridizing specifically to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex nucleic acid mixture (*e.g.*, total cellular DNA or RNA).

The phrase "hybridizing substantially to" refers to complementary hybridization between a probe nucleic acid molecule and a target nucleic acid molecule and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired hybridization.

"Stringent hybridization conditions" and "stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and Northern blot analysis are both sequence- and environment-dependent. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, part I chapter 2, Elsevier, New York, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. Typically, under "stringent conditions" a probe will hybridize specifically to its target subsequence, but to no other sequences.

The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent hybridization conditions for Southern or Northern Blot analysis of complementary nucleic acids having more than about 100 complementary residues is overnight hybridization in 50% formamide with 1 mg of heparin at 42°C. An example of highly stringent wash conditions is 15 minutes in 0.1x SSC, SM NaCl at 65°C. An example of stringent wash conditions is 15 minutes in 0.2X SSC buffer at 65°C (*See* Sambrook et al., eds (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York for a description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example of medium stringency wash conditions for a duplex of more than about 100 nucleotides, is 15 minutes in 1X SSC at 45°C. An example of low stringency wash for a duplex of more than about 100 nucleotides, is 15 minutes in 4-6X SSC at 40°C. For short probes (*e.g.*, about 10 to 50 nucleotides), stringent conditions typically involve salt concentrations of less than about 1M Na^+ ion, typically about 0.01 to 1M Na^+ ion concentration (or other salts) at pH 7.0-8.3, and the temperature is typically at least about 30°C. Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide. In general, a signal to noise ratio of 2-fold (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

The following are examples of hybridization and wash conditions that can be used to clone homologous nucleotide sequences that are substantially identical to reference nucleotide sequences of the present invention: a probe nucleotide sequence preferably hybridizes to a target nucleotide sequence in 7% sodium dodecyl sulfate (SDS), 0.5M $NaPO_4$, 1mM EDTA at 50°C followed by washing in 2X SSC, 0.1% SDS at 50°C; more preferably, a probe and target sequence hybridize in 7% sodium dodecyl sulfate (SDS), 0.5M $NaPO_4$, 1mM EDTA at 50°C followed by

washing in 1X SSC, 0.1% SDS at 50°C; more preferably, a probe and target sequence hybridize in 7% sodium dodecyl sulfate (SDS), 0.5M $NaPO_4$, 1mM EDTA at 50°C followed by washing in 0.5X SSC, 0.1% SDS at 50°C; more preferably, a probe and target sequence hybridize in 7% sodium dodecyl sulfate (SDS), 0.5M $NaPO_4$, 1mM EDTA at 50°C followed by washing in 0.1X SSC, 0.1% SDS at 50°C; more preferably, a probe and target sequence hybridize in 7% sodium dodecyl sulfate (SDS), 0.5M $NaPO_4$, 1mM EDTA at 50°C followed by washing in 0.1X SSC, 0.1% SDS at 65°C.

A further indication that two nucleic acid sequences are substantially identical is that proteins encoded by the nucleic acids are substantially identical, share an overall three-dimensional structure, are biologically functional equivalents, or are immunologically cross-reactive. These terms are defined further under the heading "Polypeptides" herein below. Nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially identical if the corresponding proteins are substantially identical. This can occur, for example, when two nucleotide sequences are significantly degenerate as permitted by the genetic code.

The term "conservatively substituted variants" refers to nucleic acid sequences having degenerate codon substitutions wherein the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer et al. (1991) *Nuc Acids Res* 19:5081; Ohtsuka et al. (1985) *J Biol Chem* 260:2605-2608; Rossolini et al. (1994) *Mol Cell Probes* 8:91-98).

The term "subsequence" refers to a sequence of nucleic acids that comprises a part of a longer nucleic acid sequence. An exemplary subsequence is a probe, described herein above, or a primer. The term "primer" as used herein refers to a contiguous sequence comprising about 8 or more deoxyribonucleotides or ribonucleotides, preferably 10-20 nucleotides, and more preferably 20-30 nucleotides of a selected nucleic acid molecule. The primers of the invention encompass oligonucleotides of sufficient length and appropriate sequence so as to provide initiation of polymerization on a nucleic acid molecule of the present invention.

The term "elongated sequence" refers to an addition of nucleotides (or other analogous molecules) incorporated into the nucleic acid. For example, a polymerase (*e.g.*, a DNA polymerase) can add sequences at the 3' terminus of the nucleic acid molecule. In addition, the nucleotide sequence can be combined with other DNA sequences, such as promoters, promoter regions, enhancers, polyadenylation signals, intronic sequences, additional restriction enzyme sites, multiple cloning sites, and other coding segments.

The term "complementary sequences", as used herein, indicates two nucleotide sequences that comprise antiparallel nucleotide sequences capable of pairing with one another upon formation of hydrogen bonds between base pairs. As used herein, the term "complementary sequences" means nucleotide sequences which are substantially complementary, as can be assessed by the same nucleotide comparison set forth above, or is defined as being capable of hybridizing to the nucleic acid segment in question under relatively stringent conditions such as those described herein. A particular example of a complementary nucleic acid segment is an antisense oligonucleotide.

The term "gene" refers broadly to any segment of DNA associated with a biological function. A gene encompasses sequences including but not limited to a coding sequence, a promoter region, a cis-regulatory sequence, a non-expressed DNA segment that is a specific recognition sequence for regulatory proteins, a non-expressed DNA segment that contributes to gene expression, a DNA segment designed to have desired parameters, or combinations thereof. A gene can be obtained by a variety of methods, including cloning from a biological sample, synthesis based on known or predicted sequence information, and recombinant derivation of an existing sequence.

The term "gene expression" generally refers to the cellular processes by which a biologically active polypeptide is produced from a DNA sequence.

The present invention also encompasses chimeric genes comprising the disclosed nuclear receptor sequences. The term "chimeric gene", as used herein, refers to a promoter region operatively linked to a nuclear

receptor coding sequence, a nucleotide sequence producing an antisense RNA molecule, a RNA molecule having tertiary structure, such as a hairpin structure, or a double-stranded RNA molecule.

The term “operatively linked”, as used herein, refers to a promoter region that is connected to a nucleotide sequence in such a way that the transcription of that nucleotide sequence is controlled and regulated by that promoter region. Techniques for operatively linking a promoter region to a nucleotide sequence are known in the art.

The terms "heterologous gene", “heterologous DNA sequence”, "heterologous nucleotide sequence”, "exogenous nucleic acid molecule”, or “exogenous DNA segment”, as used herein, each refer to a sequence that originates from a source foreign to an intended host cell or, if from the same source, is modified from its original form. Thus, a heterologous gene in a host cell includes a gene that is endogenous to the particular host cell but has been modified, for example by mutagenesis or by isolation from native cis-regulatory sequences. The terms also include non-naturally occurring multiple copies of a naturally occurring nucleotide sequence. Thus, the terms refer to a DNA segment that is foreign or heterologous to the cell, or homologous to the cell but in a position within the host cell nucleic acid wherein the element is not ordinarily found.

The term “transcription factor” generally refers to a protein that modulates gene expression by interaction with the cis-regulatory element and cellular components for transcription, including RNA Polymerase, Transcription Associated Factors (TAFs), chromatin-remodeling proteins, and any other relevant protein that impacts gene transcription.

The present invention further includes vectors comprising the disclosed nuclear sequences, including plasmids, cosmids, and viral vectors. The term “vector”, as used herein refers to a DNA molecule having sequences that enable its replication in a compatible host cell. A vector also includes nucleotide sequences to permit ligation of nucleotide sequences within the vector, wherein such nucleotide sequences are also replicated in a compatible host cell. A vector can also mediate recombinant production of a

nuclear receptor polypeptide, as described further herein below. A preferred host cell is a bacterial cell, an insect cell, or a plant cell.

Nucleic acids of the present invention can be cloned, synthesized, recombinantly altered, mutagenized, or combinations thereof. Standard recombinant DNA and molecular cloning techniques used to isolate nucleic acids are known in the art. Exemplary, non-limiting methods are described by Sambrook et al., eds (1989) Molecular Cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; by Silhavy et al. (1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; by Ausubel et al. (1992) Current Protocols in Molecular Biology, John Wylie and Sons, Inc., New York, New York; and by Glover, ed (1985) DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, United Kingdom. Site-specific mutagenesis to create base pair changes, deletions, or small insertions are also known in the art as exemplified by publications. *See e.g.*, Adelman et al. (1983) *DNA* 2:183; Sambrook et al., eds (1989) Molecular Cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Sequences detected by methods of the invention can be detected, subcloned, sequenced, and further evaluated by any measure known in the art using any method usually applied to the detection of a specific DNA sequence including but not limited to dideoxy sequencing, PCR, oligomer restriction (Saiki et al. (1985) *Bio/Technology* 3:1008-1012), allele-specific oligonucleotide (ASO) probe analysis (Conner et al. (1983) *Proc Natl Acad Sci USA* 80:278), and oligonucleotide ligation assays (OLAs) (Landgren et al. (1988) *Science* 241:1007). *See* also Landgren et al. (1988) *Science* 242:229-237.

I.B. Polypeptides

The polypeptides provided by the present invention include the isolated polypeptides set forth as SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26; polypeptides substantially identical to SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26; nuclear receptor polypeptide fragments (preferably biologically functional fragments, *e.g.* the domains described herein), fusion

proteins comprising the disclosed nuclear receptor amino acid sequences, biologically functional analogs, and polypeptides that cross-react with an antibody that specifically recognizes a disclosed nuclear receptor polypeptide.

The term “isolated”, as used in the context of a polypeptide, indicates that the polypeptide exists apart from its native environment and is not a product of nature. An isolated polypeptide can exist in a purified form or can exist in a non-native environment such as, for example, in a transgenic host cell.

The term “purified”, when applied to a polypeptide, denotes that the polypeptide is essentially free of other cellular components with which it is associated in the natural state. Preferably, a polypeptide is a homogeneous solid or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A polypeptide that is the predominant species present in a preparation is substantially purified. The term "purified" denotes that a polypeptide gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the polypeptide is at least about 50% pure, more preferably at least about 85% pure, and most preferably at least about 99% pure.

The term “substantially identical” in the context of two or more polypeptide sequences is measured as polypeptide sequences having about 35%, or 45%, or preferably from 45-55%, or more preferably 55-65% of identical or functionally equivalent amino acids. Even more preferably, two or more “substantially identical” polypeptide sequences will have about 70%, or even more preferably about 80%, still more preferably about 90%, still more preferably about 95%, and most preferably about 99% identical or functionally equivalent amino acids. Percent “identity” and methods for determining identity are defined herein below under the heading “Nucleotide and Amino Acid Sequence Comparisons”.

Substantially identical polypeptides also encompass two or more polypeptides sharing a conserved three-dimensional structure.

Computational methods can be used to compare structural representations, and structural models can be generated and easily tuned to identify similarities around important active sites or ligand binding sites. *See* Henikoff et al. (2000) *Electrophoresis* 21(9):1700-1706; Huang et al. (2000) *Pac Symp Biocomput* 230-241; Saqi et al. (1999) *Bioinformatics* 15(6):521-522; and Barton (1998) *Acta Crystallogr D Biol Crystallogr* 54:1139-1146.

The term "functionally equivalent" in the context of amino acid sequences is known in the art and is based on the relative similarity of the amino acid side-chain substituents. *See* Henikoff & Henikoff (2000) *Adv Protein Chem* 54:73-97. Relevant factors for consideration include side-chain hydrophobicity, hydrophilicity, charge, and size. For example, arginine, lysine, and histidine are all positively charged residues; that alanine, glycine, and serine are all of similar size; and that phenylalanine, tryptophan, and tyrosine all have a generally similar shape. By this analysis, described further herein below, arginine, lysine, and histidine; alanine, glycine, and serine; and phenylalanine, tryptophan, and tyrosine; are defined herein as biologically functional equivalents.

In making biologically functional equivalent amino acid substitutions, the hydropathic index of amino acids can be considered. Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics, these are: isoleucine (+ 4.5); valine (+ 4.2); leucine (+ 3.8); phenylalanine (+ 2.8); cysteine (+ 2.5); methionine (+ 1.9); alanine (+ 1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte et al. (1982) *J Mol Biol* 157:105). It is known that certain amino acids can be substituted for other amino acids having a similar hydropathic index or score and still retain a similar biological activity. In making changes based upon the hydropathic index, the substitution of amino acids whose

hydropathic indices are within ±2 of the original value is preferred, those that are within ±1 of the original value are particularly preferred, and those within ±0.5 of the original value are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Patent No. 4,554,101 states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, *e.g.*, with a biological property of the protein. It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent protein.

As detailed in U.S. Patent No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+ 3.0); lysine (+ 3.0); aspartate (+ 3.0±1); glutamate (+ 3.0±1); serine (+ 0.3); asparagine (+ 0.2); glutamine (+ 0.2); glycine (0); threonine (-0.4); proline (-0.5±1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within ±2 of the original value is preferred, those that are within ±1 of the original value are particularly preferred, and those within ±0.5 of the original value are even more particularly preferred.

The present invention also encompasses nuclear receptor polypeptide fragments or functional portions of a nuclear receptor polypeptide. Such functional portion need not comprise all or substantially all of the amino acid sequence of a native nuclear receptor gene product. The term "functional" includes any biological activity or feature of nuclear receptor, including immunogenicity.

The present invention also includes longer sequences of a nuclear receptor polypeptide, or portion thereof. For example, one or more amino acids can be added to the N-terminus or C-terminus of a nuclear receptor

polypeptide. Fusion proteins comprising nuclear receptor polypeptide sequences are also provided within the scope of the present invention. Methods of preparing such proteins are known in the art.

The present invention also encompasses functional analogs of a nuclear receptor polypeptide. Functional analogs share at least one biological function with a nuclear receptor polypeptide. An exemplary function is immunogenicity. In the context of amino acid sequence, biologically functional analogs, as used herein, are peptides in which certain, but not most or all, of the amino acids can be substituted. Functional analogs can be created at the level of the corresponding nucleic acid molecule, altering such sequence to encode desired amino acid changes. In one embodiment, changes can be introduced to improve a biological function of the polypeptide, *e.g.*, to improve the antigenicity of the polypeptide. In another embodiment, a nuclear receptor polypeptide sequence is varied so as to assess the activity of a mutant nuclear receptor polypeptide.

The present invention also encompasses recombinant production of the disclosed nuclear receptor polypeptides. Briefly, a nucleic acid sequence encoding a nuclear receptor polypeptide, or portion thereof, is cloned into an expression cassette, the cassette is introduced into a host organism, where it is recombinantly produced.

The term "expression cassette" as used herein means a DNA sequence capable of directing expression of a particular nucleotide sequence in an appropriate host cell, comprising a promoter operatively linked to the nucleotide sequence of interest which is operatively linked to termination signals. It also typically comprises sequences required for proper translation of the nucleotide sequence. The expression cassette comprising the nucleotide sequence of interest can be chimeric. The expression cassette can also be one that is naturally occurring but has been obtained in a recombinant form useful for heterologous expression.

The expression of the nucleotide sequence in the expression cassette can be under the control of a constitutive promoter or an inducible promoter

that initiates transcription only when the host cell is exposed to some particular external stimulus. Exemplary promoters include Simian virus 40 early promoter, a long terminal repeat promoter from retrovirus, an action promoter, a heat shock promoter, and a metallothien protein. In the case of a multicellular organism, the promoter and promoter region can direct expression to a particular tissue or organ or stage of development. Suitable expression vectors which can be used include, but are not limited to, the following vectors or their derivatives: viruses such as vaccinia virus or adenovirus, baculovirus vectors, yeast vectors, bacteriophage vectors (*e.g.*, lambda phage), plasmid and cosmid DNA vectors, and transposon-mediated transformation vectors.

The term "host cell", as used herein, refers to a cell into which a heterologous nucleic acid molecule has been introduced. Transformed cells, tissues, or organisms are understood to encompass not only the end product of a transformation process, but also transgenic progeny thereof.

A host cell strain can be chosen which modulates the expression of the inserted sequences, or modifies and processes the gene product in the specific fashion desired. For example, different host cells have characteristic and specific mechanisms for the translational and post-transactional processing and modification (*e.g.*, glycosylation, phosphorylation of proteins). Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. Expression in a bacterial system can be used to produce a non-glycosylated core protein product. Expression in yeast will produce a glycosylated product. Expression in insect cells can be used to ensure "native" glycosylation of a heterologous protein.

Expression constructs are transfected into a host cell by any standard method, including electroporation, calcium phosphate precipitation, DEAE-Dextran transfection, liposome-mediated transfection, transposon-mediated transformation and infection using a retrovirus. The nuclear receptor-encoding nucleotide sequence carried in the expression construct can be

stably integrated into the genome of the host or it can be present as an extrachromosomal molecule.

Isolated polypeptides and recombinantly produced polypeptides can be purified and characterized using a variety of standard techniques that are known to the skilled artisan. *See e.g.*, Ausubel *et al.* (1992) Current Protocols in Molecular Biology, John Wylie & Sons, Inc., New York, New York; Bodanszky, et al. (1976) Peptide Synthesis, John Wiley and Sons, Second Edition, New York, New York; and Zimmer et al. (1993) Peptides, pp. 393–394, ESCOM Science Publishers, B. V.

I.C. Nucleotide and Amino Acid Sequence Comparisons

The terms "identical" or percent "identity" in the context of two or more nucleotide or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms disclosed herein or by visual inspection.

The term “substantially identical” in regards to a nucleotide or polypeptide sequence means that a particular sequence varies from the sequence of a naturally occurring sequence by one or more deletions, substitutions, or additions, the net effect of which is to retain at least some of biological activity of the natural gene, gene product, or sequence. Such sequences include “mutant” sequences, or sequences wherein the biological activity is altered to some degree but retains at least some of the original biological activity. The term “naturally occurring”, as used herein, is used to describe a composition that can be found in nature as distinct from being artificially produced by man. For example, a protein or nucleotide sequence present in an organism, which can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory, is naturally occurring.

For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a

computer program, subsequence coordinates are designated if necessary, and sequence algorithm program parameters are selected. The sequence comparison algorithm then calculates the percent sequence identity for the designated test sequence(s) relative to the reference sequence, based on the selected program parameters.

Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman (1981) *Adv Appl Math* 2:482, by the homology alignment algorithm of Needleman & Wunsch (1970) *J Mol Biol* 48:443, by the search for similarity method of Pearson & Lipman (1988) *Proc Natl Acad Sci USA* 85:2444-2448, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, Madison, Wisconsin), or by visual inspection. *See* generally, Ausubel *et al.* (1992) Current Protocols in Molecular Biology, John Wylie & Sons, Inc., New York, New York.

A preferred algorithm for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al. (1990) *J Mol Biol* 215: 403-410. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold. These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits

in each direction are halted when the cumulative alignment score falls off by the quantity X from its maximum achieved value, the cumulative score goes to zero or below due to the accumulation of one or more negative-scoring residue alignments, or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength W=11, an expectation E=10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix. *See* Henikoff & Henikoff (1989) *Proc Natl Acad Sci USA* 89:10915.

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences. *See e.g.*, Karlin & Altschul (1993) *Proc Natl Acad Sci USA* 90:5873-5887. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a test nucleic acid sequence is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid sequence to the reference nucleic acid sequence is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

I.D. Antibodies

Also provided is an antibody that specifically binds an insect nuclear receptor polypeptide of the present invention. The term “antibody” indicates an immunoglobulin protein, or functional portion thereof, including a polyclonal antibody, a monoclonal antibody, a chimeric antibody, a single chain antibody, Fab fragments, and a Fab expression library. “Functional portion” refers to the part of the protein that binds a molecule of interest. In a preferred embodiment, an antibody of the invention is a monoclonal antibody. Techniques for preparing and characterizing antibodies are known in the art. *See e.g.*, Harlow & Lane (1988) Antibodies: A Laboratory Manual,

Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. A monoclonal antibody of the present invention can be readily prepared through use of well-known techniques such as the hybridoma techniques exemplified in U.S. Patent No 4,196,265 and the phage-displayed techniques disclosed in U.S. Patent No. 5,260,203.

The phrase "specifically (or selectively) binds to an antibody", or "specifically (or selectively) immunoreactive with", when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in a heterogeneous population of proteins and other biological materials. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not show significant binding to other proteins present in the sample. Specific binding to an antibody under such conditions can require an antibody that is selected based on its specificity for a particular protein. For example, antibodies raised to a protein with an amino acid sequence encoded by any of the nucleic acid sequences of the invention can be selected to obtain antibodies specifically immunoreactive with that protein and not with unrelated proteins.

The use of a molecular cloning approach to generate antibodies, particularly monoclonal antibodies, and more particularly single chain monoclonal antibodies, are also provided. The production of single chain antibodies has been described in the art. *See e.g.*, U.S. Patent No. 5,260,203. For this approach, combinatorial immunoglobulin phagemid libraries are prepared from RNA isolated from the spleen of the immunized animal, and phagemids expressing appropriate antibodies are selected by panning on tissue that expresses the polypeptide. The advantages of this approach over conventional hybridoma techniques are that approximately 10^4 times as many antibodies can be produced and screened in a single round, and that new specificities are generated by heavy (H) and light (L) chain combinations in a single chain, which further increases the chance of finding appropriate antibodies. Thus, an antibody of the present invention, or a “derivative” of an antibody of the present invention, pertains to a single polypeptide chain binding molecule which has binding specificity and affinity

substantially identical to the binding specificity and affinity of the light and heavy chain aggregate variable region of an antibody described herein.

The term "immunochemical reaction", as used herein, refers to any of a variety of immunoassay formats used to detect antibodies specifically bound to a particular protein, including but not limited to competitive and non-competitive assay systems using techniques such as radioimmunoassays, ELISA (enzyme linked immunosorbent assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitation reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using colloidal gold, enzyme or radioisotope labels), Western blot analysis, precipitation reactions, agglutination assays (*e.g.*, gel agglutination assays, hemagglutination assays), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc. *See* Harlow & Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York for a description of immunoassay formats and conditions.

I.E. Transgenic Organisms

It is also within the scope of the present invention to prepare a transgenic organism to express a transgene comprising nucleic acid sequences of the present invention. The term "transgenic organism", indicates an organism comprising a germline insertion of a heterologous nucleic acid. A transgenic organism can be an animal or a plant. Transgenic organisms of the present invention are understood to encompass not only the end product of a transformation method, but also transgenic progeny thereof.

The term "transgene", as used herein indicates a heterologous nucleic acid molecule that has been transformed into a host cell. For intended use in the creation of a transgenic organism, the transgene can include genomic sequences of the host organism at a selected locus or site of transgene integration to mediate a homologous recombination event. A transgene further comprises nucleic acid sequences of interest, for example a targeted

modification of the gene residing within the locus, a reporter gene, or a expression cassette, each defined herein above.

II. Nuclear Receptors

II.A. Conserved Features

The steroid/nuclear receptor superfamily comprises soluble receptor proteins that function as ligand-inducible transcription factors. Nuclear receptor polypeptides are characterized by the presence of five or six evolutionarily conserved domains: A/B, C, D, E and F (Evans (1988) *Science* 240:889-895; Laudet et al. (1992) *EMBO J* 11:1003-1013). Various functions are ascribed to each domain, *e.g.*, "A/B" refers to the transactivation domain, "C" refers to the DNA binding domain, "D" refers to the hinge/linker domain, "E" refers to the ligand binding domain, and "F" refers to the variable C-terminal domain that is present in some receptor polypeptides. In general, the domains can be modular in that the function of an individual domain is preserved in the context of a chimeric protein. *See e.g.*, Green & Chambon (1987) *Nature* 325:75-78; Green et al. (1988) *EMBO J* 7:3037-3044; Giguere et al. (1987) *Nature* 330:624-629.

The "A/B" region is of variable size and poorly conserved. In some cases, the A/B region has a transcriptional activation function. *See* Ptashne (1988) *Nature* 335:683-689; Hadzic et al. (1995) *Mol Cell Biol* 15:4507-4517; Pakdel et al. (1993) *Mol Endocrinol* 7:1408-1417; Thompson et al. (1989) *Proc Natl Acad Sci USA* 86:3493-3498. Amino acids that confer a transcriptional activation function can facilitate repeated transcription initiation events leading to greater levels of gene expression from a target gene.

The "C" (DNA binding) domain is a highly conserved region of approximately 198 nucleotides that encode an approximately 66 amino acid and polypeptide that comprises two Cys_2-Cys_2 zinc finger DNA binding motifs (Danielsen et al. (1989) *Cell* 57:1131-1138; Green et al. (1988) *EMBO J* 7:3037-3044; Umesono & Evans (1989) *Cell* 57:1139-1146). The DNA-binding domain also facilitates receptor dimerization (Perlmann et al. (1993) *Genes Dev* 7:1411-1422; Zechel et al. (1994) *EMBO J* 13:1414-1424; Mader

et al. (1993) *EMBO J* 12:5029-5041; Hard et al. (1990) *Science* 249:397-404). CH4 The term "core DNA binding domain" is defined as the 66 amino acid sequence that generally begins with a conserved CYS residue and ends with conserved residues GLU-MET. *See e.g.*, Rastinejad (1998) in Freedman, ed, Molecular Biology of Steroid and Nuclear Hormone Receptors, pp. 107, Birkhäuser, Boston, Massachusetts. The DNA binding domain is also characterized by a conserved three-dimensional fold when backbone atoms are compared (Rastinejad et al. (1995) *Nature* 375:203-211).

The affinity of the DNA binding domain for a DNA recognition site can be influenced by residues that are N-terminal or C-terminal to the core DNA-binding domain. *See e.g.*, Ueda et al. (1992) *Mol Cell Biol* 12:5667-5672; Rastinejad (1998) in Freedman, ed, Molecular Biology of Steroid and Nuclear Hormone Receptors, pp. 103-131, Birkhäuser, Boston, Massachusetts and references cited therein. In particular, the term "P Box" refers to sequences that are adjacent to the C-terminal end of the DNA binding domain that facilitate binding to the 5' end, generally an A/T-rich sequence of a target response element.

The "D" (hinge/linker) domain is located between the DNA binding domain and the ligand binding domain and can contribute to the strength and specificity of DNA-binding. *See e.g.*, Ueda et al. (1992) *Mol Cell Biol* 12:5667-5672; Rastinejad (1998) in Freedman, ed, Molecular Biology of Steroid and Nuclear Hormone Receptors, pp. 103-131, Birkhäuser, Boston, Massachusetts and references cited therein. In some cases, an extended DNA binding site that include sequences within the hinge domain can enable monomeric binding (*e.g.*, Ueda et al., (1992) *Mol Cell Biol* 112:5667-5672; Wilson et al. (1992) *Science* 256:106-110). The hinge region is also implicated in conferring flexibility between the ligand and DNA binding domains (Bourguet et al. (1995) *Nature* 375:377-382; Wagner et al. (1995) *Nature* 378:690-697).

The "E" (ligand binding) domain, as used herein, is interchangeably referred to as the "ligand binding domain" or the "hormone binding domain".

The ligand binding domain comprises a hydrophobic pocket that enables regulation of nuclear receptor activity by small chemical ligands. Unlike the DNA binding domain, the ligand binding domain is not clearly delineated by amino acid sequence. However, the general position of the ligand binding domain is conserved at the carboxyl end of the protein. The ligand binding domain is further characterized by a conserved tertiary structure (Wurtz et al. (1996) *Nat Struct Biol* 3:87-94).

The functional ligand binding domain within the carboxyl region of a nuclear receptor polypeptide is operationally defined as the amino acids required for high affinity binding of any ligand and can be determined according to methods known in the art. *See e.g.*, Rusconi & Yamamoto (1987) *EMBO J* 6:1309-1315; Zhang et al. (1996a) *Mol Endocrinol* 10:24-34; Lanz & Rusconi (1994) *Endocrinology* 135:2183-2194; Xu et al. (1996) *J Biol Chem* 271:21430-21438; Zhang et al. (1996b) *J Biol Chem* 271:14825-14833).

The ligand binding domain also mediates receptor dimerization. *See* Simons in Freedman, ed, Molecular Biology of Steroid and Nuclear Hormone Receptors, pp. 35-104, Birkhäuser, Boston, Massachusetts, and references cited therein. A series of heptad repeats in the ligand binding domain, specifically helix 10 with some contribution from helix 9, forms the main dimer interface (Forman et al. (1989) *Mol Endocrinol* 3:1610-1626; Forman & Samuels (1990) *Mol Endocrinol* 4:1293-1301; Bourguet et al. (1995b) *Nature* 375:377-382; Wurtz et al. (1996) *Nat Struct Biol* 3:87-94).

II.B. Identification of Novel *Drosophila* Nuclear Receptors

The present invention provides novel *Drosophila* nuclear receptor nucleic acid and polypeptide sequences. Preferably, a *Drosophila* nuclear receptor nucleic acid molecule of the present invention comprises the sequence set forth as any one of the SEQ ID NOs:1, 5, 9, 13, 17, 19, 21, 23, and 25; or a nucleic acid molecule that is substantially identical to any one of SEQ ID NOs:1, 5, 9, 13, 17, 19, 21, 23, and 25. Also preferably, a nuclear receptor polypeptide of the present invention comprises an amino acid sequence set forth as any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24,

and 26; or a polypeptide that is substantially identical to any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26.

To identify new *Drosophila* proteins, a database of predicted proteins (referred to herein as "the GeneMark database") was assembled using the GeneMark program (Borodovsky & McIninch (1993) *Computers & Chemistry* 17:123-133) and template 50 kb genomic sequence scaffolds generated by Celera Corporation (Rockville Maryland). A profile Hidden Markov Model (HMM) was built using *Drosophila* and *C. elegans* sequences as described in Example 1. The profile HMM was used to query predicted protein databases, including the GeneMark database and a predicted protein database generated by Celera using an alternative protein prediction program (referred to herein as "the Celera database").

Three new nuclear receptor sequences were identified (SEQ ID NOs:2, 6, and 10) in the GeneMark database, and three similar sequences were identified in the Celera database (SEQ ID NOs:4, 8, and 12) (Figure 1). A fourth nuclear receptor, which was designated DHR4 based on its close homology to *Tenbrio* and *Manduca* THR4 sequences, was variably predicted based on the sequences of both databases as well as other genomic clones (Accession No. AL035245). In contrast to the predicted cDNA and protein sequences, the DHR4 nucleotide sequence disclosed herein (SEQ ID NO:13) comprises an isolated DHR4 cDNA, which encodes a DHR4 protein (SEQ ID NO:14) that is different than the predicted protein sequences noted herein above.

Nuclear receptor cDNAs that encode the novel nuclear receptors identified in the GeneMark database were predicted by using the predicted receptor polypeptides to perform a TBLASTN search the genomic sequence. Nucleotide sequences retrieved by this search were assembled as the predicted nuclear receptor cDNAs. The corresponding genes were subsequently cloned as described in Example 2.

The novel *Drosophila* nuclear receptor sequences were named according to the most closely related nuclear receptor as shown in Table 2.

Table 2. BLAST Analysis of New Nuclear Receptors

	reference	(ACCESSION)	(bits)			
2	DERR	human ERRB2, complete cds (AF094517)	266	6e-70	158/375 (42%)	202/375 (53%)
6	DFAX1	C.elegans FAX-1, complete cds (AF176087)	118	3e-25	57/104 (54%)	66/104 (62%)
10	DFAX2	C.elegans FAX-1, complete cds (AF176087)	150	4e-35	72/88 (81%)	72/88 (81%)
14	DHR4	D.melanogaster BAC clone, predicted cds (AL035245)	605	e-171	358/731(48%)	358/731(48%)
		D.melanogaster genomic scaffold, predicted cds (AE003422)	506	e-142	271/331(81%)	271/331(81%)
		Bombyx GRF, complete cds (AF124981)	367	e-100	194/308 (62%)	234/308 (74%)

II. C. Identification of New *Heliothis* Nuclear Receptors

To identify new nuclear receptors in a pest insect, *Heliothis virescens* (hereinafter "*Heliothis*"), a cDNA library derived from *Heliothis* transcripts was screened using a mixture of labeled *Drosophila* nuclear receptor sequences as probe, as described in Example 3. Additional *Heliothis* nuclear receptor sequences were obtained by PCR using degenerate primers designed according to *Drosophila* nuclear receptor sequences, as described in Example 4. *Heliothis* nuclear receptor fragments derived from both methods were assembled by recognition of overlapping sequence. New *Heliothis* nuclear receptors were named based on the most closely related insect nuclear receptor (Table 3 and Figures 2-4). *Heliothis* FTZ-F1 is most similar to the β isoform encoded by *Drosophila FTZ-F1*, βFTZ-F1, in that it lacks an extended N-terminal domain characteristic of the αFTZ-F1 isoform.

Table 3. BLAST Analysis of *Heliothis* Nuclear Receptors

SEQ ID NO.	inventor's reference	*Drosophila* homologue (ACCESSION)	Score (bits)	E value	Identities	Positives
18	FTZ-F1	AAF49231	380	e-104	182/236 (77%)	202/236 (85%)
20	E75	P17671	631	e-180	312/391 (79%)	343/391 (86%)
22	USP	X52591	337	2e-91	186/413 (45%)	241/413 (58%)
24	HNF4	P49866	311	4e-84	154/216 (71%)	177/216 (81%)
26	HR3	M90806	261	5e-69	145/219 (66%)	160/219 (72%)

III. Functional Analysis of *Drosophila melanogaster* Nuclear Receptors

Many insect pests inflict plant damage by the feeding activity during larval stages. Therefore, functional analyses to assess phenotypes associated with modulation of a nuclear receptor during larval development can be used to identify candidate insecticide targets. The present invention discloses nuclear receptors whose regulation is relevant to larval viability. Thus, modulators that alter nuclear receptor activity in a manner analogous to the changes resulting from genetic manipulations described herein below, can be useful as insecticide compositions.

III.A. Loss-of-Function Analyses

RNA-mediated interference (RNAi) is a recently discovered method to determine gene function in a number of organisms, wherein double-stranded RNA (dsRNA) directs gene-specific, post-transcriptional silencing. *See e.g.*, Kuwabara & Olson (2000) *Parasitol Today* 16(8):347-349; Bass (2000) *Cell* 101(3):235-238; Hunter (2000) *Curr Biol* 10(4):R137-140; Bosher & Labouesse (2000) *Nat Cell Biol* 2(2):E31-36; Sharp (1999) Genes Dev 13(2):139-141. The double-stranded RNA molecule can be synthesized *in vitro* and then introduced into the organism by injection or other methods. Alternatively, a heritable transgene exhibiting dyad symmetry can provide a transcript that folds as a hairpin structure. Methods for examining gene functions using dsRNAi in *Drosophila* are disclosed in Example 5 and further in Kennerdell & Carthew (2000) *Nat Biotech* 18(8):896-898; Lam & Thummel (2000) *Curr Biol* 10(16):957-963; Misquitta & Paterson (1999) *Proc Natl Acad Sci USA* 96 (4):1451-1456.

The present invention discloses RNA-mediated interference of *Drosophila* nuclear receptors DERR (SEQ ID NO:2), DFAX1 (SEQ ID NO:6), DFAX2 (SEQ ID NO:10), DHR4 (SEQ ID NO:14), DSF (GenBank Accession No. AF10667, SEQ ID NO:16), FTZ-F1 (GenBank Accession M98387), EGON (GenBank Accession No. D43634), and DHR3 (GenBank Accession No. M90806) (Figure 5). Double-stranded RNA complementary to each nuclear receptor sequence was synthesized *in vitro* and injected into early

Drosophila embryos, as described in Example 5. Development of injected embryos was assessed by scoring: (a) morphological criteria using a light microscope (Campos-Ortega & Hartenstein (1985) The Embryonic Development of *Drosophila melanogaster*, Springer-Verlag, Berlin), (b) embryo hatching to become a larvae, (c) puparium formation, and (d) eclosion of the pupae as an adult fly, as indicated in Table 4 herein below. Buffer-injected embryos were injected and monitored in parallel as a control. The percentage of embryos injected with dsRNA that survive to the adult stage is depicted in Figure 5.

Table 4. Results of dsRNAi Analysis

dsRNA	# eggs injected	# embryos showing morphological development	# larvae	# pupae	# adults	% embryonic lethality	% larval lethality	% pupal lethality	viable
DHR3	212	181	15	0	0	92%	8%	0	0
DHR4	405	352	7	7	4	98%	0	<1%	1%
EGON	166	127	93	3	2	27%	71%	<1%	2%
FTZ-F1	373	235	13	10	9	94%	1%	<1%	4%
DSF	103	88	13	9	99	85%	5%	0	10%
DERR	304	270	223	59	55	17%	61%	1%	20%
DFAX1	333	273	147	124	123	46%	8%	<1%	46%
DFAX2	468	418	251	199	192	40%	12%	2%	46%
buffer	941	806	580	500	433	28%	10%	8%	54%

Percentage lethality at each developmental stage was calculated as follows:

% embryonic lethality =[(# embryos showing morphological development) – (# larvae)] / # embryos showing morphological development

% larval lethality = [(# larvae) – (# pupae)] / # embryos showing morphological development

% pupal lethality = [(# pupae) – (# adults)] / # embryos showing morphological development

% viable = # adults / # embryos showing morphological development

Injection of double-stranded *DHR3*, *FTZ-F1*, and *EGON* RNA conferred significant embryonic lethality (92% and 94% respectively). Embryos injected with double-stranded *EGON* RNA also showed lethality during larval stages (71%). The observed lethal phases resulting from loss of *DHR3*, *FTZ-F1*, and *EGON*, as determined using dsRNAi, are consistent with published loss-of-function phenotypes for *DHR3* (Carney et al. (1997) *Proc Natl Acad Sci USA* 94(22):12024-12029), *FTZ-F1* (Yu et al. (1997) *Nature* 385:552-555; Guichet et al. (1997) *Nature* 385:548-552), and *EGON* (Dittrich et al. (1997) *Development* 124(13):2515-2525).

Embryos injected with double-stranded *DHR4* or *DSF* RNA showed significant embryonic lethality (98% and 85%, respectively). *DHR4* and *DSF* can also be required during larval development as many genes are essential at multiple developmental stages. Injection of double-stranded *DERR* RNA resulted in lethality predominantly during larval stages (61%). By contrast, injection of double-stranded DFAX1 or DFAX2 RNA showed some lethality during embryonic stages, although not substantially different than buffer-injected control animals.

Lethality resulting from loss of nuclear receptor function is predicted to be mimicked by provision of an antagonist substance that specifically binds a given receptor. Nuclear receptor antagonists can be identified by methods known in the art and as further disclosed in the section entitled Identification of Insect Nuclear Receptor Modulators, herein below. The essentiality of nuclear receptors DHR4, DSF, and DERR, disclosed herein for the first time, identifies the utility of antagonists that block or mitigate the activity of DHR4, DSF, and DERR as insecticides.

III.B. Gain-of-Function Analyses

Ectopic expression systems have been used to elucidate gene function when classical loss-of-function genetics is not straightforward. For example, heat-induced expression of *spaghetti squash*, which encodes the nonmuscle myosin II regulatory light chain, can effectively rescue the early lethality of *spaghetti squash* mutants, facilitating the analysis of phenotypes later in development (Edwards & Kiehart (1996) *Development* 122:1499).

Similarly, dominant phenotypes generated by overexpressing a gene of interest have been used to address post-embryonic gene functions, particularly in cases where gene mutation results in embryonic lethality. *See e.g.*, Lam et al. (1999) *Dev Biol* 212(1):204-216; Woodard et al. (1994) *Cell* 79(4):607-615).

Transgenic methods for ectopic expression in *Drosophila* utilize promoters that drive either constitutive or regulated expression of the gene of interest. Constructs designed for ectopic expression can be prepared in a transformation vector, and are introduced into the fly genome by germ line transformation. A transgenic line is established, and ectopic expression of the gene of interest can be analyzed in a wild type or mutant genetic background.

In one embodiment, a heat shock promoter can be used to temporally regulate gene expression (Lis et al. (1983) *Cell* 35:403; Struhl (1985) *Nature* 318:677; Schneuwly et al. (1987) *Nature* 325:816). Using this approach, the level of ectopic gene expression can be easily modulated by altering the temperature and/or duration of the heat treatment.

Overexpression of nuclear receptors can reveal the role of an activated nuclear receptor. Provision of ligand and/or apo-receptor (a receptor not bound by ligand) favors formation of the liganded receptor. Similarly, provision of excess nuclear receptor overexpression can also lead to an excess of active receptor. *See e.g.*, Tsai et al. (1998) in Wilson et al., eds, Williams Textbook of Endocrinology, pp. 55-94, W.B. Saunders Company, Philadelphia, Pennsylvania, and references cited therein. To create this situation *in vivo*, a nuclear receptor can be overexpressd using a heterologous transgene. This strategy enables a functional assessment of orphan nuclear receptors, wherein a ligand has not yet been identified. A phenotype observed following nuclear receptor overexpression is predicted to also be generated by abnormally elevated levels of endogenous ligand or by administration of a nuclear receptor agonist.

The present invention discloses overexpression of nuclear receptors during *Drosophila* larval development. Transgenic *Drosophila* lines were

employed that carry heat inducible nuclear receptor transgenes, as described in Example 6. During *Drosophila* development, the first larval instar begins with hatching of the embryo, and culminates with the first larval molt at approximately 24 hours after hatching (25°C). Transgenic larvae were briefly heat treated (1 hr at 37°C) at the beginning of larval development (0-2 hrs +/- 15 min) or alternatively at the end of larval development (20-22 hrs +/- 15 min), immediately prior to molting. Control experiments omitted heat treatment. The genotype from which the transgenic lines are derived, w^{1118}, were treated in parallel experiments as an additional control. The developmental progress of larvae was monitored at 24 hours following heat treatment and at puparium formation.

Heat-induced expression of DHR38, DHR39, and E75A at the end of the first instar also resulted in larval lethality (Table 5, Figure 6). By contrast, transgenic larvae that were not heat-treated were substantially viable, demonstrating that the presence of the transgene, in the absence of induced nuclear receptor expression, is not responsible for the observed lethality. Further, transgenic larvae that were heat treated at the beginning of first instar larval development, w^{1118} larvae that were heat treated either at the beginning or end of larval development, and non-heat-treated w^{1118} larvae were also substantially viable. Thus, the lethality observed in transgenic larvae that were heat treated just prior to molting (20-22 hr) is attributable to induced nuclear receptor expression, and is not a consequence of heat treatment alone.

Table 5. Percent Viability of Heat-Treated Larvae

larvae	heat treatment		
	none	0-2 hr	20-22 hr
control	88%	76%	88%
hs-DHR38	92%	86%	16%
hs-DHR39	90%	78%	12%
hs-E75A	84%	72%	26%

Examination of lethal larvae that were heat treated to induce expression of DHR39 revealed the presence of multiple mouthhooks (28%). This phenotype is consistent with a defect in epidermal molting, whereby mouthhooks are normally expelled (Perrimon et al. (1985) *Genetics* 11:23-41; Demerec (1965) Biology of *Drosophila*, Hafner Publishing, New York, New York.

Lethality resulting from overexpression of DHR38, DHR39, and E75A is predicted to be mimicked by provision of agonists that bind these receptors. Nuclear receptor agonists can be identified by methods known in the art and as further disclosed in the section entitled Identification of Insect Nuclear Receptor Modulators, herein below. Prior to the disclosure of the subject application, the larval lethal phenotypes conferred by overexpression of DHR38, DHR39, and E75A was unknown. The phenotypic characterization of nuclear receptor modulation during larval development, disclosed herein, identifies the utility of nuclear receptor agonists that activate DHR38, DHR39, E75A, and homologues of the indicated nuclear receptors as insecticides.

Gain-of-function phenotypes of new *Drosophila* nuclear receptors DERR (SEQ ID NO:2), DFAX1 (SEQ ID NO:6), and DFAX2 (SEQ ID NO:10) disclosed herein can be addressed using ectopic expression techniques in *Drosophila* that are known in the art. The present invention provides nucleotide sequences (SEQ ID NOs:1, 5, and 9) encoding such receptors that can be used to construct vectors for ectopic expression.

IV. Recombinant Expression of Insect Nuclear Receptors

For recombinant production of a protein of the invention in a host organism, a nucleotide sequence encoding the protein is inserted into an expression cassette designed for the chosen host and introduced into the host where it is recombinantly produced. The choice of the specific regulatory sequences such as promoter, signal sequence, 5' and 3' untranslated sequence, and enhancer appropriate for the chosen host is within the level of ordinary skill in the art. The resultant molecule, containing

the individual elements linking in the proper reading frame, is inserted into a vector capable of being transformed into the host cell.

Expression constructs can be transfected into a host cell by a standard method suitable for the selected host, including electroporation, calcium phosphate precipitation, DEAE-Dextran transfection, liposome-mediated transfection, infection using a retrovirus, transposon-mediated transfer, and particle bombardment techniques. The expression cassette sequence carried in the expression construct can be stably integrated into the genome of the host or it can be present as an extrachromosomal molecule.

Suitable expression vectors and methods for recombinant production of proteins are known for host organisms such as *E. coli*, yeast, and insect cells. *See e.g.*, Lucknow & Summers (1988) *Bio/Technol* 6:47. Representative methods for recombinant production of an insect nuclear receptor in *E. coli* are disclosed in Example 7.

Additional suitable expression vectors are baculovirus expression vectors, *e.g.*, those derived from the genome of *Autographica californica* nuclear polyhedrosis virus (AcMNPV). A preferred baculovirus/insect system is PVL1392/PVL1393 used to transfect *Spodoptera frugiperda* (SF9) cells in the presence of linear *Autographica californica* baculovirus DNA (Pharmingen of San Diego, California). The resulting virus is used to infect HighFive *Trichoplusia ni* cells (Invitrogen Corporation of Carlsbad, California). Representative methods for recombinant production of an insect nuclear receptor in insect cells are disclosed in Example 8.

Recombinantly produced proteins can be isolated and purified using a variety of standard techniques. The actual techniques used varies depending upon the host organism used, whether the protein is designed for secretion, and other such factors. Such techniques are known to the skilled artisan. *See* Ausubel *et al.* (1992) Current Protocols in Molecular Biology, John Wylie and Sons, Inc., New York, New York.

The present invention further encompasses recombinant expression of the disclosed insect nuclear receptors, or portion thereof, in plants, as described further herein below under the section entitled Transgenic Plants.

V. Production of Insect Nuclear Receptor Antibodies

In another aspect, the present invention provides a method of producing an antibody immunoreactive with an insect nuclear receptor polypeptide, the method comprising recombinantly or synthetically producing an insect nuclear receptor polypeptide, or portion thereof, to be used as an antigen. The insect nuclear receptor polypeptide is formulated so that it is can be used as an effective immunogen. An animal is immunized with the formulated insect nuclear receptor polypeptide to generate an immune response in the animal. The immune response is characterized by the production of antibodies that can be collected from the blood serum of the animal. Optionally, cells producing an insect nuclear receptor antibody can be fused with myeloma cells, whereby a monoclonal antibody can be selected. Exemplary methods for producing a monoclonal antibody that recognizes an insect nuclear receptor protein are described in Example 4. Preferred embodiments of the method use a polypeptide set forth as any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26.

The present invention also encompasses antibodies and cell lines that produce monoclonal antibodies as described herein.

The foregoing antibodies can be used in methods known in the art relating to the localization and activity of the insect nuclear receptor polypeptide sequences of the invention, *e.g.*, for cloning of insect nuclear receptor nucleic acids, immunopurification of insect nuclear receptor polypeptides, imaging insect nuclear receptor polypeptides in a biological sample, and measuring levels thereof in appropriate biological samples.

VI. Methods for Detecting an Insect Receptor Nucleic Acid

In another aspect of the invention, a method is provided for detecting a nucleic acid molecule that encodes an insect nuclear receptor polypeptide. Such methods can be used to detect insect nuclear receptor gene variants

and related resistance gene sequences. The disclosed methods facilitate genotyping, cloning, gene mapping, and gene expression studies.

The nucleic acids of the present invention can be used to clone genes and genomic DNA comprising the disclosed sequences. Alternatively, the nucleic acids of the present invention can be used to clone genes and genomic DNA of related sequences, preferably nuclear receptor genes in pest insects and nematodes. Using the nucleic acid sequences disclosed herein, such methods are known to one skilled in the art. *See*, for example, Sambrook et al., eds (1989) Molecular Cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. Representative methods are also disclosed in Examples 3 and 4. Preferably, the nucleic acids used for this method comprise sequences set forth as any one of SEQ ID NOs:1, 5, 9, 13, 17, 19, 21, 23, and 25.

In one embodiment, genetic assays based on nucleic acid molecules of the present invention can be used to screen for genetic variants by a number of PCR-based techniques, including single-strand conformation polymorphism (SSCP) analysis (Orita et al. (1989) *Proc Natl Acad Sci USA* 86(8):2766-2770), SSCP/heteroduplex analysis, enzyme mismatch cleavage, direct sequence analysis of amplified exons (Kestila et al. (1998) *Mol Cell* 1(4):575-582; Yuan et al. (1999) *Hum Mutat* 14(5):440-446), allele-specific hybridization (Stoneking et al. (1991) *Am J Hum Genet* 48(2):370-82), and restriction analysis of amplified genomic DNA containing the specific mutation. Automated methods can also be applied to large-scale characterization of single nucleotide polymorphisms (Brookes (1999) *Gene* 234(2):177-186; Wang et al. (1998) *Science* 280(5366):1077-1082). Preferred detection methods are non-electrophoretic, including, for example, the TAQMAN™ allelic discrimination assay, PCR-OLA, molecular beacons, padlock probes, and well fluorescence. *See* Landegren et al. (1998) *Genome Res* 8:769-776.

VII. Methods for Detecting an Insect Nuclear Receptor Polypeptide

In another aspect of the invention, a method is provided for detecting a level of insect nuclear receptor polypeptide using an antibody that

specifically recognizes an insect nuclear receptor polypeptide, or portion thereof. In a preferred embodiment, biological samples from an experimental subject and a control subject are obtained, and insect nuclear receptor polypeptide is detected in each sample by immunochemical reaction with the insect nuclear receptor antibody. More preferably, the antibody recognizes amino acids of any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26; and is prepared according to a method of the present invention for producing such an antibody.

In one embodiment, an insect nuclear receptor antibody is used to screen a biological sample for the presence of an insect nuclear receptor polypeptide. A biological sample to be screened can be a biological fluid such as extracellular or intracellular fluid, or a cell or tissue extract or homogenate. A biological sample can also be an isolated cell (*e.g.*, in culture) or a collection of cells such as in a tissue sample or histology sample. A tissue sample can be suspended in a liquid medium or fixed onto a solid support such as a microscope slide. In accordance with a screening assay method, a biological sample is exposed to an antibody immunoreactive with an insect nuclear receptor polypeptide whose presence is being assayed, and the formation of antibody-polypeptide complexes is detected. Techniques for detecting such antibody-antigen conjugates or complexes are known in the art and include but are not limited to centrifugation, affinity chromatography and the like, and binding of a labeled secondary antibody to the antibody-candidate receptor complex.

A modulator that shows specific binding to an insect modulator can also be used to detect an insect nuclear receptor. Representative techniques for assaying specific binding include are described herein above under the heading "Identification of Insect Nuclear Receptor Modulators".

The disclosed methods for detecting an insect nuclear receptor polypeptide can be useful to determine altered levels of gene expression that are associated with particular conditions such as enhanced tolerance to insecticides that target a particular insect nuclear receptor polypeptide.

VIII. Identification of Nuclear Receptor Modulators

The present invention further discloses a method for identifying a compound that modulates an insect nuclear receptor. As used herein, the terms "candidate substance" and "candidate compound" are used interchangeably and refer to a substance that is believed to interact with another moiety, wherein a biological activity is modulated. For example, a representative candidate compound is believed to interact with an insect nuclear receptor polypeptide, or fragment thereof, and can be subsequently evaluated for such an interaction. Exemplary candidate compounds that can be investigated using the methods of the present invention include, but are not restricted to, viral epitopes, peptides, enzymes, enzyme substrates, co-factors, lectins, sugars, oligonucleotides or nucleic acids, oligosaccharides, proteins, chemical compounds, small molecules, and antibodies. A candidate compound to be tested can be a purified molecule, a homogenous sample, or a mixture of molecules or compounds.

As used herein, the term "modulate" means an increase, decrease, or other alteration of any or all chemical and biological activities or properties of a wild-type insect nuclear receptor polypeptide, preferably an insect nuclear receptor polypeptide of any one of the even-numbered SEQ ID NOs:2-34. Preferably, an insect nuclear receptor modulator is an agonist of an insect nuclear receptor protein activity. As used herein, the term "agonist" means a substance that synergizes or potentiates the biological activity of a functional insect nuclear receptor protein. As used herein, the term "antagonist" refers to a substance that blocks or mitigates the biological activity of an insect nuclear receptor polypeptide.

In accordance with the present invention there is also provided a rapid and high throughput screening method that relies on the methods described above. This screening method comprises separately contacting each compound with a plurality of substantially identical target molecules. In such a screening method the plurality of target molecules preferably comprises more than about 10^4 samples, or more preferably comprises more than about 5×10^4 samples. In an alternative high-throughput strategy, each target molecule can be contacted with a plurality of candidate compounds.

The disclosed methods can also be used to identify a modulator that interacts with an insect nuclear receptor ligand binding domain. Such assays can employ a target molecule comprising the full-length nuclear receptor polypeptide. Alternatively, an isolated ligand binding domain can be recombinantly produced for use in the assay. *See* Coffer et al. (1996) *J Steroid Biochem Mol Biol* 58:467-477; Tetel et al. (1997) *Mol Endocrinol* 11:1114-1128; Rochel et al. (1997) *Biochem Biophys Res Commun* 230:293-296). Optionally, additional sequences, such as receptor, GST, or polyhistidine sequences, can be fused to the amino terminal of the ligand binding domain to stabilize the conformation of the recombinantly expressed ligand binding domain. *See e.g.*, Simental et al. (1991) *J Biol Chem* 266:510-518; Nemoto et al. (1992) *J Steroid Biochem Mol Biol* 42:803-812; Cooper et al. (1996) *J Steroid Biochem Mol Biol* 57:251-257; Eul et al. (1989) *EMBO J* 8:83-90; Lin et al. (1991) *Mol Endocrinol* 5:485-492; Wagner et al. (1995) *Nature* 378:690-697; Dallery et al. (1993) *Biochem* 32:12428-12436; Lupisella et al. (1995) *J Biol Chem* 270:24884-24890; Rochel et al. (1997) *Biochem Biophys Res Commun* 230:293-296; Leng et al. (1995) *Mol Cell Biol* 15:255-263.

In one embodiment, the disclosed methods for identifying modulators of insect nuclear receptors are performed using nuclear receptor sequences set forth as any one of even-numbered 2-34. In particular, the loss-of-function larval lethality that is observed when DHR4, DERR, or DSF function is disrupted (Figure 5) suggests that antagonists of DHR4, DERR, or DSF can be useful as insecticides. The larval lethal phenotype that is observed when DHR38, DHR39, or E75A is overexpressed in *Drosophila* (Figure 6) suggest that agonists of DHR38, DHR39, or E75A can be useful as insecticides.

The disclosed methods for identifying modulators of insect nuclear receptors can be performed using nucleic acid sequences derived from a pest organism. The nuclear receptor sequences disclosed herein provide methods for identifying homologous sequences in pest species. Such techniques are well know to those in the art. *See* for example, Sambrook et

al., eds (1989) Molecular Cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, and Examples 3 and 4 herein below.

In a preferred embodiment, the disclosed methods for identifying modulators employ a *Heliothis* βFTZ-F1 (SEQ ID NO:18), *Heliothis* E75 (SEQ ID NO:20), or *Heliothis* USP (SEQ ID NO:22) polypeptide. The larval lethal phenotype that is observed when E75A is overexpressed in *Drosophila* (Figure 5), suggest that agonists of E75A can be useful as insecticides. Genetic data in *Drosophila* shows that loss-of-function mutation in any one of *USP, βFTZ-F1,* or *E75A*, also confers larval lethality (Yu et al. (1997) *Nature* 385:552-555; Johnson & Garza (1998) *Ann Dros Res Conf* 39:430A), suggesting that antagonists of USP, βFTZ-F1, and E75A can also be useful as insecticides.

Representative methods for identification of a substance that binds and thereby modulates an insect nuclear receptor are disclosed herein below. The term "binding" refers to an affinity between two molecules, for example, a ligand and a receptor. As used herein, "binding" means a preferential binding of one molecule for another in a mixture of molecules. The binding of the molecules can be considered specific if the binding affinity is about 1×10^4 M^{-1} to about 1×10^6 M^{-1} or greater. Binding of two molecules also encompasses a quality or state of mutual action such that an activity of one protein or compound on another protein is inhibitory (in the case of an antagonist) or enhancing (in the case of an agonist). To demonstrate saturable binding of a candidate compound, identified by any such method, to a nuclear receptor ligand binding domain, Scatchard analysis can be carried out as described, for example, by Mak et al. (1989) *J Biol Chem* 264:21613:21618.

VIII.A. Protein Binding Assays

Several techniques can be used to detect interactions between a protein and a chemical ligand without employing an *in vivo* ligand. Representative methods include, but are not limited to, Fluorescence Correlation Spectroscopy, Surface-Enhanced Laser Desorption/Ionization Time-Of-flight Spectroscopy, and Biacore technology, as described herein

below. These methods are amenable to automated, high-throughput screening.

Fluorescence Correlation Spectroscopy (FCS) measures the average diffusion rate of a fluorescent molecule within a small sample volume (Madge et al. (1972) *Phys Rev Lett* 29:705-708; Maiti et al. (1997) *Proc Natl Acad Sci USA* 94:11753-11757). The sample size can be as low as 10^3 fluorescent molecules and the sample volume as low as the cytoplasm of a single bacterium. The diffusion rate is a function of the mass of the molecule and decreases as the mass increases. FCS can therefore be applied to polypeptide-ligand interaction analysis by measuring the change in mass and therefore in diffusion rate of a molecule upon binding. In a typical experiment, the target to be analyzed is expressed as a recombinant polypeptide with a sequence tag, such as a poly-histidine sequence, inserted at the N-terminus or C-terminus. The expression takes place in *E. coli*, yeast or mammalian cells. The polypeptide is purified using chromatographic methods. For example, the poly-histidine tag can be used to bind the expressed polypeptide to a metal chelate column such as Ni^{2+} chelated on iminodiacetic acid agarose. The polypeptide is then labeled with a fluorescent tag such as carboxytetramethylrhodamine or BODIPY™ (Molecular Probes of Eugene, Oregon). The polypeptide is then exposed in solution to the potential ligand, and its diffusion rate is determined by FCS using instrumentation available from Carl Zeiss, Inc. (Thornwood, New York). Ligand binding is determined by changes in the diffusion rate of the polypeptide.

Surface-Enhanced Laser Desorption/Ionization (SELDI) was developed by Hutchens & Yip (1993) *Rapid Commun Mass Spectrom* 7:576-580. When coupled to a time-of-flight mass spectrometer (TOF), SELDI provides a technique to rapidly analyze molecules retained on a chip. It can be applied to ligand-protein interaction analysis by covalently binding the target protein, or portion thereof, on the chip and analyzing by MS the small molecules that bind to this protein (Worrall et al. (1998) *Anal Biochem* 70:750-756). In a typical experiment, the target to be analyzed is expressed

as described for FCS. The purified protein is then used in the assay without further preparation. It is bound to the SELDI chip either by utilizing the poly-histidine tag or by other interaction such as ion exchange or hydrophobic interaction. The chip thus prepared is then exposed to the potential ligand via, for example, a delivery system able to pipet the ligands in a sequential manner (autosampler). The chip is then washed in solutions of increasing stringency, for example a series of washes with buffer solutions containing an increasing ionic strength. After each wash, the bound material is analyzed by submitting the chip to SELDI-TOF. Ligands that specifically bind the target are identified by the stringency of the wash needed to elute them.

Biacore relies on changes in the refractive index at the surface layer upon binding of a ligand to a target polypeptide immobilized on the layer. In this system, a collection of small ligands is injected sequentially in a 2-5 microliter cell, wherein the target polypeptide is immobilized within the cell. Binding is detected by surface plasmon resonance (SPR) by recording laser light refracting from the surface. In general, the refractive index change for a given change of mass concentration at the surface layer is practically the same for all proteins and peptides, allowing a single method to be applicable for any protein (Liedberg et al. (1983) *Sensors Actuators* 4:299-304; Malmquist (1993) *Nature* 361:186-187). In a typical experiment, the target to be analyzed is expressed as described for FCS. The purified protein is then used in the assay without further preparation. It is bound to the Biacore chip either by utilizing the poly-histidine tag or by other interaction such as ion exchange or hydrophobic interaction. The chip thus prepared is then exposed to the potential ligand via the delivery system incorporated in the instruments sold by Biacore (Uppsala, Sweden) to pipet the ligands in a sequential manner (autosampler). The SPR signal on the chip is recorded and changes in the refractive index indicate an interaction between the immobilized target and the ligand. Analysis of the signal kinetics of on rate and off rate allows the discrimination between non-specific and specific interaction.

VIII.B. Peptide Interaction Assays

Methods for displaying diverse peptide libraries enable rapid library construction, amplification, and selection of ligands directed against a target molecule. *See* Lowman (1997) *Annu Rev Biophys Biomol Struct* 26:401-424; Sidhu (2000) *Curr Opin Biotech* 11(6):610-616; and U.S. Patent No. 5,510,240. Assays can also be employed that select peptides capable of disrupting the interaction between a nuclear receptor and a requisite co-factor, as described by Hall et al. (2000) *Mol Enodcrinol* 14(12):2010-2023; Northrop et al. (2000) *Mol Endocrinol* 14(5)605-622; International Publication No. WO 00/37077, herein incorporated by reference.

VIII.C. Transcriptional Assays

The present invention also provides methods for identifying modulators of insect nuclear receptor transcriptional activation. One strategy employs an expression system comprising: (1) an insect nuclear receptor comprising a functional ligand binding domain of an insect nuclear receptor, (2) a target gene expression cassette comprising a response element regulated by the chimeric nuclear receptor operatively linked to a reporter gene, and (3) a test compound. Methods for constructing a chimeric nuclear receptor gene and a target gene expression cassette are described herein below under the heading "Chimeric Receptors for Inducible Gene Expression". *See* also, Wentworth et al. (2000) *J Endocrinol* 166(3):R11-16; Yang & Chen (1999) *Cancer Res* 59(18):4519-4524, and U.S. Patent No. 4,981,784, herein incorporated by reference.

The term "reporter gene" refers to a heterologous gene encoding a product that is readily observed and/or quantitated. A reporter gene is heterologous in that it originates from a source foreign to an intended host cell or, if from the same source, is modified from its original form. Any suitable reporter and detection method can be used in accordance with the disclosed methods. Non-limiting examples of detectable reporter genes that can be operatively linked to a transcriptional regulatory region can be found in Alam and Cook (1990) *Anal Biochem* 188:245-254 and International Publication No. WO 97/47763. Preferred reporter genes for transcriptional

analyses include the *lacZ* gene (*See e.g.*, Rose and Botstein (1983) *Meth Enzymol* 101:167-180), Green Fluorescent Protein (GFP) (Cubitt et al. (1995) *Trends Biochem Sci* 20:448-455), luciferase, or chloramphenicol acetyl transferase (CAT).

An amount of reporter gene can be assayed by any method for qualitatively, or preferably quantitatively, determining presence or activity of the reporter gene product. The amount of reporter gene expression directed by each test substance is compared to an amount of reporter gene expression in the absence of a test substance. A test substance is identified as having agonist activity when there is significant increase in a level of reporter gene expression in the presence of the substance when compared to a level of reporter gene expression in the absence of the test substance. The term "significant increase", as used herein, refers to an quantified change in a measurable quality that is larger than the margin of error inherent in the measurement technique, preferably an increase by about 2-fold or greater relative to a control measurement, more preferably an increase by about 5-fold or greater, and most preferably an increase by about 10-fold or greater.

VIII.D. Rational Design

The knowledge of the structure a native nuclear receptor polypeptide provides an approach for rational pesticide design. *See e.g.* Schapira et al. (2000) *Proc Natl Acad Sci USA* 97(3):1008-1013. The structure of a nuclear receptor polypeptide can be determined by X-ray crystallography or by computational algorithms that generate three-dimensional representations. *See* Huang et al. (2000) *Pac Symp Biocomput* 230-41; Saqi et al. (1999) *Bioinformatics* 15:521-522; International Publication No. WO 99/26966, herein incorporated by reference. Alternatively, a working model of a nuclear receptor structure can be derived by homology modeling (Maalouf et al. (1998) *J Biomol Struc Dynamics* 15(5):841-851). Computer models can further predict binding of a protein structure to various substrate molecules, which can be synthesized and tested. Additional compound design techniques are described in U.S. Patent Nos. 5,834,228 and 5,872,011.

IX. Methods for Pest Control

Another aspect of the present invention is a method for pest control by modulation of insect nuclear receptor biological activity. Substances having such activity can be discovered by the methods disclosed herein and include, but are not limited to, chemical compounds, antibodies, and gene products encoded by plant transgenes.

The present invention provides methods for preventing the onset or progression of a pest infestation in a plant. The method comprises administering a modulator of a nuclear receptor set forth as any one of the even-numbered SEQ ID NOs:2-34, wherein modulation of the nuclear receptor results in organismal lethality. Preferably, the lethality occurs during larval development.

IX.A. Formulation

An insect nuclear receptor modulator of the present invention is typically formulated using acceptable vehicles, adjuvants, and carriers as desired. Representative formulations include emulsifiable concentrates, water-miscible liquids, wettable powders, water-soluble powders, oil solutions, flowable powders, aerosols, vapors, granulars, microcapsules, fumigants, ultra-low volume concentrates, fogging concentrates, vapors, impregnating materials, poison baits, and seed dressings. *See e.g.*, Perry et al. (1997) Insecticides in Agriculture and Environment: Retrospects and Prospects, pp. 7-10, Springer-Verlag, New York, New York. A formulation can be further selected based on its ability to improve insecticide properties such as storage, handling, application, effectiveness, safety to the applicator and the environment, and cost.

An insecticide formulation can further include a synergist that can enhance the activity of an insect nuclear receptor modulator of the present invention. *See* Yamamoto (1973) in Casida, ed, Pyrethrum, The Natural Insecticide, pp. 191-170, Academic Press, New York, New York; Hodgson & Tate (1976) in Wilkinson, ed, Insecticide Biochemistry and Physiology, pp. 115-148, Plenum Press, New York, New York; Wilkinson (1976a) in Tahori, ed, Proc 2nd Int Congr on Pesticides and Chemistry, Vol. 2, pp. 117-159,

Gordon & Breach, New York, New York; Wilkinson (1976b) in Metcalf & McKelvey, eds, The Future for Insecticides: Needs and Prospects, Vol. 6, pp. 191-178, Wiley, New York, New York; Casida & Quistad (1995) in Casida & Quistad, eds, Pyrethrum Flowers: Production, Chemistry, Toxicology, and Uses, pp. 258-276, Oxford University Press, New York, New York. Alternatively, synergism can be accomplished by treatment of a plant prior to application of an insect nuclear receptor modulator, or by application of a synergist at sites on a plant distinct form sites of application of an insect nuclear receptor modulator.

IX.B. *In Vivo* Assays

The insecticidal activity of a modulator of an insect nuclear receptor can be tested using standard techniques in the art, including topical application, injection, dipping, contact or residual exposure, and feeding/drinking. *See e.g.*, Perry et al. (1997) Insecticides in Agriculture and Environment: Retrospects and Prospects, pp. 12-13, Springer-Verlag, New York, New York. As one example, a formulation comprising a modulator is sprayed on a plant, insect larvae are then applied to the plant, and after an appropriate temporal duration, a degree of plant destruction by the larvae is quantitated.

IX.C. Dose and Administration

The toxicity of an insecticide to an organism can be expressed in terms of the amount of compound per unit weight of the organism required to kill 50% of the test population, also referred to as the lethal dose (LD_{50}). The LD_{50} is usually expressed in milligrams per kilogram of body weight or micrograms per insect. The lethal concentration (LC_{50}) is the concentration of a compound in an external medium that is required to kill 50% of the test population, and is expressed as the percentage or parts per million (ppm) of the active ingredient (AI) in the medium. This value can be used when the exact dose administered to an insect cannot be determined. The effectiveness of a candidate insecticidal substance can also be assayed in terms of lethal time (LC_{50}). LC_{50} represents the time required to elicit 50%

mortality of the test organisms at a specified dose or concentration and is a suitable measure for field tests.

In some instances, a rate of knockdown rather than lethality is measured as a criterion of effectiveness. In such cases the knockdown dose (KD_{50}) or the knockdown time (KT_{50}) can be used to express insecticidal activity.

The present invention also envisions the identification of insecticidal substances wherein killing or knockdown does not constitute the desired criterion. For example, useful assays can also assess non-lethal measures such as, for example, progression to developmental stages, fecundity, egg viability, attractant or repellant activity, paralysis, and anti-feeding activity.

Insect nuclear receptor modulators identified in accordance with methods of the present invention are useful for preventing or treating an insect infestation, and in some cases a nematode infestation, in a plant or animal. Prevention and treatment methods employ an effective amount of the modulator. The term "effective amount" as used herein refers to an amount effective to prevent or ameliorate infestation.

An effective amount can comprise a range of amounts. One skilled in the art can readily assess the potency and efficacy of an insect nuclear receptor modulator of the present invention and adjust the administration regimen accordingly. A modulator of insect nuclear receptor biological activity can be evaluated by a variety of techniques, for example, by using a responsive reporter gene in an transcriptional assay, by assaying interaction of insect nuclear receptor polypeptides with a monoclonal antibody, or by assaying insect viability when a modulator is administered to an insect, each technique described herein. One of ordinary skill in the art can tailor the dosages to a particular application, taking into account the particular formulation and method of administration to be used with the composition as well as the type of plant or animal, the development stage of the plant or animal, and the severity of the infestation to be treated.

IX.D. Transgenic Plants

The present invention also encompasses methods for pest control wherein an insect nuclear receptor modulator is expressed in a plant. Preferably, a nucleic acid, peptide or polypeptide encoded by a transgene in a plant modulates the activity of any of SEQ ID NOs:1-34. In one embodiment, a transgene can encode a peptide that specifically binds an insect nuclear receptor of the present invention. In another embodiment, a construct encoding an antibody that specifically binds an insect nuclear receptor of the present invention can be expressed in plants to confer insect control. *See e.g.*, U.S. Patent No. 5,686,600, the contents of which are herein fully incorporated by reference. Methods for generating a transgenic plant are known in the art and are discussed further herein below.

IX.E. Target Organisms

Insect nuclear receptor modulators discovered according to the methods disclosed herein can be used for the prevention or amelioration of a pest infestation. The term "pest" as used herein refers to any organism that damages a plant, including mature plants, seedlings, and stored grain. The term "pest" also refers to any organism that causes disease in an animal. The compositions and methods disclosed herein are envisioned to be particularly useful to prevent or to treat infestation of insect pests, including but not limited to aphids, locusts, spider mites, boll weevils, and pests that attack stored grains (*e.g.*, *Tribolium* and *Tenebrio*). The present disclosure is also relevant to methods for controlling soil nematodes and plant-parasitic nematodes such as *Melooidogyne*.

X. Chimeric Receptors for Inducible Gene Expression

Transgenic methods have enabled the generation of plants with improved traits by expression of a transgene encoding a heterologous polypeptide of interest. Ideally, the temporal profile of transgene expression can be controlled.

The present invention envisions a gene switch method that employs three or more components: (1) a nuclear receptor expression cassette, (2) a ligand that binds the polypeptide encoded by the nuclear receptor expression cassette, and (3) a target gene expression cassette that is

modulated in the presence of the encoded nuclear receptor polypeptide further bound by a ligand.

A nuclear expression receptor cassette of the present invention encodes a nuclear receptor polypeptide. In one embodiment, the nuclear receptor polypeptide is composed of a hinge region, a ligand binding domain, a DNA binding domain, and a transactivation domain. The DNA binding domain enables binding of the nuclear receptor polypeptide to a sequence-specific response element in the 5' regulatory region of a target expression cassette. The hinge domain of the receptor polypeptide resides between the DNA binding and ligand binding domains and influences the activity of the ligand binding domain. The ligand binding domain of the receptor polypeptide can bind a chemical ligand, thereby eliciting a conformational change in the receptor polypeptide that allows the transactivation domain to affect transcription of the target nucleotide sequence.

A "target expression cassette" comprises a nucleotide sequence for a 5' regulatory region operatively linked to a target nucleotide sequence, the expression of which is activated by a receptor polypeptide in the presence of a chemical ligand. The 5' regulatory region of the target gene comprises a core promoter sequence, an initiation of transcription sequence, and one or more sequence-specific response elements required for receptor binding to the target gene regulatory region. The promoter sequence can be a minimal promoter. The target expression cassette can also possess a 3' termination region (stop codon and polyadenylation sequence). The target nucleotide sequence can encode, for example, a polypeptide, an antisense RNA, or a double-stranded RNA molecule. A target sequence can be part of a target cassette transformed into a host organism, or it can be a target sequence of a native host organism gene.

The chimeric receptor polypeptides used in the present invention can have one or more domains obtained from a heterologous source. The use of chimeric receptor polypeptides has the benefit of combining domains from different sources, thus providing a receptor polypeptide activated by a choice

of chemical ligands and possessing desirable ligand binding, DNA binding and transactivation characteristics.

It is also considered a part of the present invention that the transactivation (A/B), ligand-binding (E), and DNA-binding (C) domains can be assembled in the chimeric receptor polypeptide in any functional arrangement. For example, where one subdomain of a transactivation domain is found at the N-terminal portion of a naturally-occurring receptor, a chimeric receptor polypeptide of the present invention can include a transactivation domain at the C-terminus in place of, or in addition to, a transactivation domain at the N-terminus. Chimeric receptor polypeptides as disclosed herein can also have multiple domains of the same type, for example, more than one transactivation domain per receptor polypeptide.

X.A. DNA Binding Domain of a Receptor Expression Cassette

A chimeric receptor of the present invention can comprise a DNA-binding domain from any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26. The term "DNA binding domain" as used in the context of a chimeric receptor of the present invention comprises a functional domain that shows high affinity sequence-specific DNA binding. A functional DNA binding domain will generally include the core DNA binding domain, and optionally, sequences adjacent to the core DNA binding domain that contribute to high affinity specific-specific DNA binding.

A gene switch receptor cassette preferably encodes a chimeric nuclear receptor that modulates gene expression as a monomer or dimer. *Drosophila* βFTZ-F1 and the vertebrate homologue of βFTZ-F1, SF-1, can bind to an extended half-site response element as a monomer and can further function as a transcriptional activator in this context (Ueda et al. (1992) *Mol Cell Biol* 12(12):5667-5672; Wilson et al. (1993) *Mol Cell Biol* 13(9):5794-5804). Similarly, a vertebrate homologue of E75, RevErb, represses transcription on both monomeric and dimeric binding sites (Harding & Lazar (1995) *Mol Cell Biol* 15(11):6479; Lazar & Harding (1998) in Freedman, ed, Molecular Biology of Steroid and Nuclear Hormone Receptors, pp. 269-270, Birkhäuser, Boston, Massachusetts). The

Drosophila DNA-binding domain and C-terminated extension of the core DNA binding domain is highly homologous to RevErb and shows similar DNA-binding properties (Segraves & Hogness (1990) *Genes Dev* 4:204-219; Lazar & Harding (1998) in Freedman, ed, Molecular Biology of Steroid and Nuclear Hormone Receptors, pp. 270, Birkhäuser, Boston, Massachusetts).

The *Heliothis* βFTZ-F1 and E75A nuclear receptors disclosed herein are highly conserved compared to their *Drosophila* and vertebrate homologues (Figures 3 and 4). In particular, the DNA binding domain and the C-terminal extension of the core DNA-binding domain of *Heliothis* βFTZ-F1 (SEQ ID NO:18) is substantially identical to *Drosophila* βFTZ-F1. The DNA binding domain and the C-terminal extension of the core DNA-binding domain of *Heliothis* E75 (SEQ ID NO:20) is substantially identical compared to *Drosophila* E75A. Given the similar DNA binding properties between *Drosophila* βFTZ-F1 and vertebrate SF-1, and between *Drosophila* E75A and vertebrate RevErb, *Heliothis* βFTZ-F1 and E75 likely function as monomeric and/or dimeric transcriptional regulators as well. Thus, the DNA binding domains, optionally with the extended C-terminal DNA binding sequence, can be useful in gene switch receptor expression cassettes that operate as monomeric transcriptional regulators.

The vertebrate nuclear receptor most closely related to DERR, estrogen-related receptor, also binds and activates transcription as a monomer, again utilizing an extended C-terminal DNA binding sequence (Bonnelye et al. (1997) *Mol Endocrinol* 11(7):905-916; Johnston et al. (1997) *Mol Endocrinol* 11(3):342-352). Thus, the DERR DNA binding domain, optionally also including the C-terminal extended DNA binding sequence, can be useful in gene switch receptor expression cassettes.

Additional flexibility in controlling gene expression by the present invention can be obtained by using DNA binding domains and response elements from other transcriptional activators, which include but are not limited to the bacterial LexA or yeast GAL4 proteins (Brent & Ptashne (1985) *Cell* 43: 729-736i; Sadowski et al. (1988) *Nature* 335:563-564).

An additional degree of flexibility in controlling gene expression can be obtained by using synthetic DNA binding domains and response elements. Protein engineering experiments can rationally alter the DNA binding characteristics of zinc finger domains to bind to a DNA target sequence of choice (Liu et al. (1997) *Proc Natl Acad Sci USA* 94:5525-5530; Desjarlais & Berg (1993) *Proc Natl Acad Sci USA* 90:2252-1860). For example, the use of a synthetic zinc finger binding domain allows the chimeric receptor polypeptide to recognize a target sequence of choice.

A chimeric receptor expression cassette of the present invention comprises a suitable promoter operatively linked to the coding sequence intended for expression. The expression of the nucleotide sequence in the expression cassette can be under the control of a constitutive promoter, an inducible promoter, or a tissue-specific promoter. Depending upon the host cell system utilized, any one of a number of suitable promoters can be used. Promoter selection can be based on expression profile and expression level.

X.B. Target Gene Expression Cassette

A target gene expression cassette includes a response element that is operatively linked to a target gene of interest.

In one embodiment, gene switch systems of the present invention can employ a chimeric receptor expression polypeptide having a DNA binding domain that binds a well-characterized response element, for example, the LexA and GAL4 binding sites (Brent & Ptashne (1985) *Cell* 43: 729-736; Sadowski et al. (1988) *Nature* 335:563-564).

In another embodiment, a gene switch system of the present invention can employ a chimeric receptor polypeptide comprising a DNA binding domain derived from a novel insect nuclear receptor disclosed herein (SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26). Response elements that are recognized by DNA binding domains of novel nuclear receptors can be determined according to standard methods. Briefly, *in vivo* footprinting assays can demonstrate protection of DNA sequences from chemical and enzymatic modification within living or permeabilized cells. Similarly, *in vitro* footprinting assays can protect DNA sequences from chemical or enzymatic

modification using protein extracts. Nitrocellulose filter-binding assays and gel electrophoresis mobility shift assays (EMSAs) can track the presence of radiolabeled regulatory DNA elements based on provision of candidate transcription factors. Computer analysis programs, for example TFSEARCH version 1.3 (Yutaka Akiyama: "TFSEARCH: Searching Transcription Factor Binding Sites", http://www.rwcp.or.jp/papia/), can be used to locate consensus sequences of known cis-regulatory elements within a genomic region.

In a preferred embodiment of the invention, multiple copies of the appropriate response element are placed in the 5' regulatory region, which allows multiple sites for binding of receptor polypeptide resulting in a greater degree of activation.

X.C. Ligand Binding Domain of a Receptor Expression Cassette

Preferably, the ligand binding domain of a chimeric receptor of the present invention comprises the “E” domain of any one of SEQ ID NOs:2, 6, 10, 14, 18, 20, 22, 24, and 26. For the purpose of creating a receptor expression cassette, the “E” domain is defined operationally as a sequence that is sufficient for ligand binding and can be determined by methods known in the art. A ligand binding domain of an insect nuclear receptor disclosed herein can further be modified to permit and/or optimize binding to a selected ligand. Such modifications can include point mutations and truncation.

X.D. Activation/Repression Domain of the Receptor Expression Cassette

Transactivation (A/B) domains can be defined as amino acid sequences that, when combined with the DNA binding domain in a receptor polypeptide, increase productive transcription initiation by RNA polymerases. *See*, generally, Ptashne (1988) *Nature* 335:683-689; Meshi (1995) *Plant Cell Physiol* 36:1405-1420. Different transactivation domains are known to have different degrees of effectiveness in their abilities to increase transcription initiation. In the present invention, it is desirable to use transactivation domains that have superior transactivating effectiveness in host cells in order to create a high level of target expression cassette expression in

response to the presence of chemical ligand. Representative transactivation domains include but are not limited to herpes simplex virus VP16 (Triezenberg et al. (1988) *Genes Dev* 2(6):718-729), maize C1 (Goff et al. (1991) *Genes and Dev* 5:298-309), *Arabidopsis* AP1, and maize Dof1.

As described above, the method of the present invention can be used to increase gene expression over a minimal, basal level. One of the outstanding benefits of the present method, however, is that it can also be used for decreasing or inhibiting gene expression, *e.g.*, gene repression. Controlling gene expression through repression can be accomplished using a repression domain in place of the transactivation domain. Repression domains can be defined as amino acid sequences that, when combined with the DNA binding domain in a receptor polypeptide, decrease the productive transcription initiation by RNA polymerases (Ng (2000) *Trends Biochem Sci* 25:121-126). Repression domains that can be used with the present invention to decrease expression of a target cassette include but are not limited to the repression domains of AtHD2A (Wu (2000) *Plant J* 22:19-27), Oshox1, and Oshox3 (Meijer (2000) *Mol Gen Genet* 263:12-21).

X.E. Additional Features of Gene Switch Expression Cassettes

The transgenic expression of genes can also involve modification of the relevant transgenes to achieve and optimize their expression in a selected host. Modifications can include but are not limited to cloning of open reading frames normally encoded by a single gene in separate expression cassettes, utilization of host codon preferences, adjustment of GC/AT content, inclusion of sequences adjacent to the initiating methionine codon that promote efficient translation, and removal of illegitimate splice sites.

An expression cassette can also comprise any additional sequences required or selected for the expression of the transgene. Such sequences include, but are not limited to, transcription terminators, introns, sequences that can enhance gene expression, and sequences that mediate intracellular targeting of the gene product.

X.F. Ligands for Inducible Gene Expression

A ligand that activates a chimeric receptor polypeptide and that can be used in a gene switch expression system can be identified, for example, using methods described herein above under the heading "Identification of Insect Nuclear Receptor Modulators".

XI. Transgenic Plants

The present invention envisions expression of insect nuclear receptor modulators and components of nuclear receptor gene switch expression systems in plants. Representative techniques for transforming dicotyledonous and monocotyledonous plants are described herein below.

The phrase "a plant, or parts thereof", as used herein shall mean an entire plant; and shall mean the individual parts thereof, including but not limited to seeds, leaves, stems, and roots, as well as plant tissue cultures. Transgenic plants of the present invention are understood to encompass not only the end product of a transformation method, but also transgenic progeny thereof.

Representative plants that can be used in transgenic methods disclosed herein include but are not limited to rice, wheat, barley, rye, corn, potato, carrot, sweet potato, sugar beet, bean, pea, chicory, lettuce, cabbage, cauliflower, broccoli, turnip, radish, spinach, asparagus, onion, garlic, eggplant, pepper, celery, carrot, squash, pumpkin, zucchini, cucumber, apple, pear, quince, melon, plum, cherry, peach, nectarine, apricot, strawberry, grape, raspberry, blackberry, pineapple, avocado, papaya, mango, banana, tobacco, tomato, sorghum and sugarcane.

XI.A. Promoters

For *in vivo* production of an insect nuclear receptor modulator or a chimeric receptor expression cassette in plants, exemplary constitutive promoters are derived from the CaMV 35S, rice actin, and maize ubiquitin genes. *See* Binet et al. (1991) *Plant Sci* 79:87-94, Christensen et al. (1989) *Plant Mol Biol* 12:619-632, Callis et al. (1990) *J Biol Chem* 265:12486-12493, Norris et al. (1993) *Plant Mol Biol* 21:895-906, European Patent Application Nos. 0 342 926 and 0 392 225, Taylor et al (1993) *Plant Cell Rep*

12:491-495, McElroy et al (1990) *Plant Cell* 2:163-171, McEclroy et al. (1991) *Mol Gen Genet* 231:150-160, Chibbar et al. (1993) *Plant Cell Rep* 12:506-509. Representative inducible promoters suitable for use with the present invention include the chemically inducible *PR-1* promoter, the *PR-1a* promoter, an ethanol-inducible promoter, a glucocorticoid inducible promoter, and a wound-inducible promoter. *See* Uknes et al. (1992) *Plant Cell* 4:645-656, Lebel et al. (1998) *Plant J* 16:223-233, Caddik et al. (1998) *Nat Biotechnol* 16:177-180, Aoyama & Chua (1997) *The Plant Journal* 11:605-612, Xu et al. (1993) *Plant Mol Biol* 22:573-588; Logemann et al. (1989) *Plant Cell* 1:151-158, Rohrmeier & Lehle (1993) *Plant Mol Biol* 22:783-792, Firek et al. (1993) *Plant Mol Biol* 22:129-142, and Warner et al. (1993) *Plant J* 3:191-201. Selected promoters can direct expression in specific cell types (such as leaf epidermal cells, mesophyll cells, root cortex cells) or in specific tissues or organs (roots, leaves or flowers, for example). Representative promoters that direct cell- or tissue-specific expression in plants and can be used in accordance with the present invention include but are not limited to a root-specific promoter (de Framond (1991) *FEBS* 290:103-106, U.S. Patent No. 5,466,785), a pith-preferred promoter (International Publication No. WO 93/07278), a leaf-specific promoter (Hudspeth & Grula (1989) *Plant Mol Biol* 12:579-589), and a pollen-specific promoter (International Publication No. WO 93/07278).

XI.B. Vectors

The expression cassette is cloned into a vector suitable for transformation. Suitable expression vectors which can be used include, but are not limited to, the following vectors or their derivatives: plant transformation vectors, viruses such as vaccinia virus or adenovirus, baculovirus vectors, yeast vectors, bacteriophage vectors (*e.g.*, lambda phage), plasmid and cosmid DNA vectors, and transposon-mediated transformation vectors.

Numerous vectors available for plant transformation are known to those of ordinary skill in the plant transformation arts, and the genes pertinent to this invention can be used with any such vectors. Exemplary

vectors include pCIB200, pCIB2001, pCIB10, pCIB3064, pSOG19, and pSOG35. The selection of vector will depend upon the preferred transformation technique and the target species for transformation.

Many vectors are available for transformation using *Agrobacterium tumefaciens*. These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan (1984) *Nuc Acids Res* 12:8711-8721) and pXYZ. *See* also European Patent Application No. 0 332 104, herein incorporated by reference.

Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector and consequently vectors lacking these sequences can be utilized in addition to vectors such as the ones described above which contain T-DNA sequences. Transformation techniques that do not rely on *Agrobacterium* include transformation via particle bombardment, protoplast uptake (*e.g.*, electroporation), and microinjection. The choice of vector depends largely on the preferred selection for the plant species being transformed.

For certain target species, different antibiotic or herbicide selection markers can be preferred. Selection markers used routinely in transformation include the *nptII* gene, which confers resistance to kanamycin and related antibiotics (Messing & Vierra (1982) *Gene* 19: 259-268; Bevan et al. (1983) *Nature* 304:184-187), the *bar* gene, which confers resistance to the herbicide phosphinothricin (White et al. (1990) *Nuc Acids Res* 18:1062, Spencer et al. (1990) *Theor Appl Genet* 79:625-631), the *hph* gene, which confers resistance to the antibiotic hygromycin (Blochlinger & Diggelmann (1984) *Mol Cell Biol* 4:2929-2931), and the *dhfr* gene, which confers resistance to methatrexate (Bourouis et al. (1983) *EMBO J* 2(7):1099-1104), the *EPSPS* gene, which confers resistance to glyphosate (U.S. Patent Nos. 4,940,935 and 5,188,642), and the mannose-6-phosphate isomerase gene, which provides the ability to metabolize mannose (U.S. Patent Nos. 5,767,378 and 5,994,629).

XI.C. Transformation of Dicotyledons

Transformation techniques for dicotyledons are known in the art and include *Agrobacterium*-based techniques and techniques that do not require *Agrobacterium*. Non-*Agrobacterium* techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This can be accomplished by polyethylene glycol (PEG) electroporation, particle bombardment-mediated uptake, or microinjection. Examples of these techniques are described by Paszkowski et al. (1984) *EMBO J* 3:2717-2722; Potrykus et al. (1985) *Mol Gen Genet* 199:169-177; Reich et al. (1986) *Biotechnology* 4:1001-1004; Klein et al. (1987) *Nature* 327:70-73; and U.S. Patent Nos. 4,945,050, 5,036,006, and 5,100,792. Using any of the afore-mentioned methods, the transformed cells can be regenerated to whole plants using standard techniques known in the art.

XI.D. Transformation of Monocotyledons

Transformation of most monocotyledon species has now also become routine. Preferred techniques include direct gene transfer into protoplasts using PEG or electroporation techniques, and particle bombardment into callus tissue. *See* European Patent Application Nos. 0 292 435, 0 392 225, and 0 332 581; International Publication Nos. WO 93/07278 and WO 93/21335; Gordon-Kamm et al. (1990) *Plant Cell* 2:603-618; Fromm et al. (1990) *Biotechnology* 8:833-839; Koziel et al. (1993) *Biotechnology* 11:194-200; Zhang et al. (1988) *Plant Cell Rep* 7:379-384; Shimamoto et al. (1989) *Nature* 338:274-277; Datta et al. (1990) *Biotechnology* 8:736-740; Christou et al. (1991) *Biotechnology* 9:957-962; Vasil et al. (1992) *Biotechnology* 10:667-674; Vasil et al. (1993) *Biotechnology* 11:1553-1558; and Weeks et al. (1993) *Plant Physiol* 102:1077-1084. More recently, transformation of monocotyledons using *Agrobacterium* has been described. *See* International Publication No. WO 94/00977 and U.S. Patent No. 5,591,616, both of which are incorporated herein by reference.

XII. Methods of Inducible Gene Expression

The present invention further provides a method of controlling gene expression in an organism, the method comprising: (a) transforming the

organism with a receptor expression cassette comprising a 5' regulatory region capable of promoting expression operatively linked to a receptor cassette encoding a chimeric receptor polypeptide of the invention, and a 3' terminating region; (b) transforming the organism with a target expression cassette comprising a 5' regulatory region operatively linked to a target nucleotide sequence, wherein the 5' regulatory region comprises one or more response elements that are recognized by the DNA binding domain of the chimeric receptor polypeptide; (c) expressing the chimeric receptor polypeptide in the organism; and (d) contacting the organism with a chemical ligand that binds to the ligand binding domain of the chimeric receptor polypeptide, whereby the chimeric receptor polypeptide activates expression of the target nucleotide sequence.

Methods employing a gene switch system as disclosed herein are useful for regulated expression in any organism that can express a nuclear receptor expression cassette, including both plants and animals.

In a preferred embodiment, nuclear receptor cassettes comprising disclosed nuclear receptor sequences are useful for the regulation of expression of target polypeptides in plants in the presence of appropriate chemical ligands. For example, U.S. Patent No. 5,880,333 is drawn to a method for controlling gene expression in plants comprising transforming a plant with expression cassette encoding a nuclear receptor polypeptide and a target sequence. The method is useful for controlling various traits of agronomic importance.

The gene switch system disclosed herein can also be adopted to gene therapy methods. The insect nuclear receptor components of a receptor cassette disclosed herein are particularly relevant to regulated expression in mammals based on their heterologous derivation. Utilization of insect nuclear receptor ligand binding domains, and ligands that specifically activate such nuclear receptors, will minimize the possibility of cross-reactivity with endogenous mammalian receptors.

Examples

The following Examples have been included to illustrate modes of the invention. Certain aspects of the following Examples are described in terms of techniques and procedures found or contemplated by the present co-inventors to work well in the practice of the invention. These Examples illustrate standard laboratory practices of the co-inventors. In light of the present disclosure and the general level of skill in the art, those of skill will appreciate that the following Examples are intended to be exemplary only and that numerous changes, modifications, and alterations can be employed without departing from the scope of the invention.

Example 1

Database Searches

A pileup alignment of *Drosophila* and *C. elegans* nuclear receptor DNA-binding domains was generated using the GCG program (Devereux et al. (1984) *Nuc Acids Res* 12:387-395). The pileup was used to build a profile Hidden Markov Model (HMM) according to the HMMER 2.1.1 program (available from Washington University School of Medicine, St. Louis, Missouri). HMMER 2.1.1 hmmbuild parameters were selected for maximal sensitivity for identifying complete fragments and excluding local alignments. The profile HMM was further calibrated according to the program instructions. A database (referred to herein as "the GeneMark database") was assembled by predicting proteins using the GeneMark program (Borodovsky & Mclninch (1993) *Computers & Chemistry* 17:123-133) based on 50 kb *Drosophila* genomic sequence scaffolds that had been generated at Celera Genomics, Inc. (Rockville, Maryland). The profile HMM was also used to search a database generated by Celera using alternative protein prediction programs (referred to herein as the "Celera predicted protein database").

Example 2

Isolation of *Drosophila melanogaster* Nuclear Receptor cDNAs

cDNA clones of *Drosophila* nuclear receptors were cloned by PCR using a first strand *Drosophila* cDNA pool as template. PCR primers were designed to include the predicted start and stop codons of each receptor using the primer3 application (available through Whitehead/MIT Center for Genome Research of Cambridge, Massachusetts). Amplified products were cloned in the pUNI/V5-His-TOPO vector (Invitrogen Corporation of Carlsbad, California). Cloned inserts were sequences on both strands by primer walking using an ABI PRISM® 3700 DNA Analyzer (Applied Biosystems of Foster City, California) to an accuracy of <1/10,000 nucleotide errors.

Example 3

Cloning *Heliothis virescens* Nuclear Receptors by Library Screening

A cDNA library was constructed in the Uni-ZAP XR vector Stratagene of La Jolla, California) using oligo-dT-primed transcripts from *Heliothis virescens* embryos aged 0-24 hours at 25°C. Library clones were excised from Uni-ZAP XR as pBluescript phagemids. The library was transformed into *E.coli* by electroporation and plated into 384-well plates with an average calculated density of 4-6 colonies per well. Plates were arrayed by a Q-Bot (Genetix Pharmaceuticals, Inc. of Cambridge, Massachusetts) onto a 22 x 22 cm nylon filter. The filter was probed under low stringency conditions. Representative low-stringency conditions are overnight hybridization at 55°C in Church's Buffer followed by washing in 2X SSC 0.1% SDS, twice at room temperature and once at 42°C, each wash approximately 30 min). The probe was a mixture of random primed radioactively labeled DNA binding domains (DBDs) of *Drosophila melanogaster* nuclear receptors *EcR*, *usp*, *DHR38*, *DHR39*, *βFTZ-F1* and *E75*. The DBD fragments were generated by PCR using nondegenerate primers corresponding to the indicated DBDs.

For secondary screening, *E.coli* from positive wells were spread onto agar plates, 96 colonies corresponding to each primary screen well were

picked into 384-well plates. The colonies were arrayed onto a nylon membrane and re-probed using identical conditions. Positive clones were sequenced.

Example 4

Cloning *Heliothis virescens* Nuclear Receptors by PCR

polyA$^+$ RNA was made from *Heliothis virescens* embryos aged 0-24 hours at 25°C. cDNA was prepared by reverse transcription using random primers and a Gene Amp kit (Perkin Elmer of San Jose, California). The reverse transcription reaction was allowed to proceed for 15 min at 42°C, and the reaction was stopped by incubating the reaction for 5 min at 99°C. For amplification of nuclear receptors, nested, degenerate primers designed according to the *Drosophila* βFTZ-F1 sequence (SEQ ID NOs:35-38). Primers included restriction enzyme sites to facilitate cloning into pBluescript (Stratagene of La Jolla, California). Cycling parameters for amplification using degenerate primers were as follows: initial amplification -- 2 min at 95°C; 35 cycles -- 15 sec at 95°C, 30 sec at 60°C; 7 min at 72°C; hold at 4°C; second amplification -- 2 min at 95°C; 35 cycles --15 sec at 95°C, 30 sec at 60°C, 2.5 min at 72°C; 7 min at 72°C; hold at 4°C.

Amplification of cDNA ends was performed using a MARATHON RACE kit (Clontech Laboratories, Inc. of Palo Alto, California). Gene-specific primers (SEQ ID NOs:39-42) were designed according to *Heliothis virescens* nuclear receptor sequences obtained as described herein above. cDNA libraries generated from *Heliothis virescens* adult head and/or larval gut were used as template. RACE products were cloned into the TOPO-A vector (Invitrogen Corporation of Carlsbad, California).

Example 5

Double-Stranded RNA Interference

Preparation of dsRNA for Injection. Sequences to be expressed as dsRNA were cloned into Bluescript KS(+) (Stratagene of La Jolla, California), linearized with the appropriate restriction enzymes, and transcribed *in vitro*

with the Ambion T3 and T7 MEGASCRIPT® high yield transcription kits following the manufacturer's instructions (Ambion Inc. of Austin, Texas). Transcripts were annealed in injection buffer (0.1mM $NaPO_4$ pH 7.8, 5mM KCl) after heating to 85°C and cooling to room temperature over a 1- to 24-hr period. All annealed transcripts were analyzed on agarose gels with DNA markers to confirm the size of the annealed RNA and quantitated as described previously (Fire et al. (1998) *Nature* 391(6669):806-811). Injected RNA was not gel-purified. Injection of 0.1 nl of a 0.1- to 1.0-mg/ml solution of a 1-kb dsRNA corresponds to roughly 10^7 molecules/injection.

Injection of *Drosophila melanogaster* Embryos. Fly cages were set up using 2- to 4-day flies. Agar-grape juice plates were replaced every hour to synchronize the egg collection for 1-2 days. The eggs were collected over a 30- to 60-min period for subsequent injection. The eggs were washed into a nylon mesh basket with tap water. The chorion was removed by brief soaking in a dilute bleach solution. Eggs were positioned on a glass slide such that each egg was in a same orientation. Double-stranded RNA was injected into middle of each egg using an Eppendorf transjector (Eppendorf Scientific, Inc. of Westbury, New York). Following injection, slides were stored in a moist chamber to prevent dessication of the embryos. Embryos were monitored for development and transferred as first instar larvae to vials containing *Drosophila* medium. Methods for rearing *Drosophila* staging and common genetic techniques can be found, for example, in Roberts (1986) *Drosophila melanogaster*, A Practical Approach, IRL Press, Washington, DC; Ashburner (1989a) *Drosophila*: A Laboratory Handbook, Cold Spring Harbor Laboratory Press, New York, New York; Ashburner (1989b) *Drosophila*: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York, New York; Goldstein & Fyrberg, eds (1994) in Methods in Cell Biology, Vol. 44, Academic Press, San Diego, California.

Example 6

Overexpression of Nuclear Receptors in *Drosophila melanogaster*

The following *Drosophila melanogaster* lines were used for over-expression analysis:

w^{1118}

w^{1118}; +/SM5; P[*hs-E75A* w^{+}]/TM3

w^{1118}; P[*hs-E75A-H* w^{+}]/CyO; Dr/TM3

w^{1118}; P[*hs-DHR39-6* w^{+}]

w^{1118}; P[*hs-DHR39-3* w^{+}]

yw; P[*hs-DHR38* w^{+}]-II

All lines were acquired from the public *Drosophila melanogaster* Stock Center of Bloomington, Indiana (http://flystocks.bio.indiana.edu) or from Dr. Carl Thummel of Howard Hughes Medical Institute, Salt Lake City, Utah. *Drosophila melanogaster* genotypes are indicated according to standard nomenclature in the field. *See* Lindsley & Zimm (1992) The Genome of *Drosophila melanogaster*, Academic Press, San Diego, California.

Embryos were collected on grape juice/agar plates at 25°C. First instar larvae were staged at embryo hatching. Collections of first instar larvae aged 0-2 hours +/- 15 minutes post-hatching were allowed to develop to the desired stage at 25°C.

Heat treatments were performed by placing staged larvae on grape juice/agar plates in a 37°C warm room for 2 hours. Larvae were then transferred to vials containing *Drosophila* medium and returned to 25°C for post-heat treatment recovery. The duration of heat treatment was determined empirically to result in minimal lethality of control lines. Viability was scored at 24 hours post heat treatment by observation of larval movement and/or at four days post heat treatment by counting the number of pupae. All heat treatment experiments were performed using collections of 50 larvae per genotype.

Example 7

Recombinant Production of in *E. coli*

A cDNA clone of the present invention is subcloned into an appropriate expression vector and transformed into *E. coli* using the manufacturer's conditions. Specific examples include plasmids such as pBluescript (Stratagene of La Jolla, California), pFLAG (International Biotechnologies, Inc. of New Haven, Connecticut), and pTrcHis (Invitrogen Corporation of Carlsbad, California). *E. coli* are cultured, expression of the recombinant protein is confirmed, and recombinant protein is isolated using standard techniques.

Example 8

Recombinant Production of a Nuclear Receptor in Insect Cells

Baculovirus vectors, which are derived from the genome of AcNPV virus, are designed to provide high levels of expression of cDNA in the *Spodoptera frugiperda* (SF9) line of insect cells (ATCC CRL# 1711). Recombinant baculovirus expressing the cDNA of the present invention is produced by the following standard methods (Invitrogen MaxBac Manual, Invitrogen Corporation of Carlsbad, California): cDNA constructs are ligated into the *polyhedrin* gene in any one of a variety of baculovirus transfer vectors, including the pAC360 and the BleBAc vector (Invitrogen Corporation of Carlsbad, California). Recombinant baculoviruses are generated by homologous recombination following co-transfection of the baculovirus transfer vector and linearized AcNPV genomic DNA (Kitts (1990) *Nucleic Acid Res* 18:5667) into SF9 cells. Recombinant pAC360 viruses are identified by the absence of inclusion bodies in infected cells and recombinant pBlueBac viruses are identified on the basis of β-galactosidase expression (Summers & Smith, *Texas Agriculture Exp Station Bulletin* No. 1555).

A cDNA encoding an entire open reading frame for gene is inserted into the *BamH* I site of pBlueBacII (Invitrogen Corporation of Carlsbad,

California). Constructs in the positive orientation, identified by sequence analysis, are used to transfect SF9 cells in the presence of linear AcNPV wild type DNA. The recombinant insect nuclear receptor is present in the cytoplasm of infected cells. The recombinant insect nuclear receptor is extracted from infected cells by hypotonic or detergent lysis.

Example 9

In vitro Binding Assays

Recombinant protein can be obtained, for example, according to the approach described in Example 7 or 8 herein above. The protein is immobilized on chips appropriate for ligand binding assays. The protein immobilized on the chip is exposed to a candidate substance according to methods known in the art. While the sample compound is in contact with the immobilized protein, measurements capable of detecting protein-ligand interactions are conducted. Measurement techniques include, but are not limited to, SEDLI, Biacore, and FCS, as described above. Substances that bind the protein are readily discovered using this approach and are subjected to further characterization.

References

The references listed below as well as all references cited in the specification are incorporated herein by reference to the extent that they supplement, explain, provide a background for or teach methodology, techniques and/or compositions employed herein.

Adelman et al. (1983) *DNA* 2:183.

Alam and Cook (1990) *Anal Biochem* 188:245-254.

Altschul et al. (1990) *J Mol Biol* 215: 403-410.

Aoyama & Chua (1997) *The Plant Journal* 11:605-612.

Ashburner (1989a) *Drosophila*: A Laboratory Handbook, Cold Spring Harbor Laboratory Press, New York, New York.

Ashburner (1989b) *Drosophila*: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York, New York.

Ausubel et al. (1992) Current Protocols in Molecular Biology, John Wylie and Sons, Inc., New York, New York.

Barton (1998) *Acta Crystallogr D Biol Crystallogr* 54:1139-1146.

Bass (2000) *Cell* 101(3):235-238.

Batzer et al. (1991) *Nuc Acids Res* 19:5081.

Bevan (1984) *Nuc Acids Res* 12:8711-8721.

Bevan et al. (1983) *Nature* 304:184-187.

Binet et al. (1991) *Plant Sci* 79:87-94.

Blochlinger & Diggelmann (1984) *Mol Cell Biol* 4:2929-2931.

Bodanszky, et al. (1976) Peptide Synthesis, John Wiley and Sons, Second Edition, New York, New York.

Bonnelye et al. (1997) *Mol Endocrinol* 11(7):905-916.

Borodovsky & McIninch (1993) *Computers & Chemistry* 17:123-133.

Bosher & Labouesse (2000) *Nat Cell Biol* 2(2):E31-36.

Bourguet et al. (1995) *Nature* 375:377-382.

Bourouis et al. (1983) *EMBO J* 2(7):1099-1104.

Brent & Ptashne (1985) *Cell* 43: 729-736.

Brookes (1999) *Gene* 234(2):177-186.

Caddik et al. (1998) *Nat Biotechnol* 16:177-180.

Callis et al. (1990) *J Biol Chem* 265:12486-12493.

Campos-Ortega & Hartenstein (1985) The Embryonic Development of *Drosophila melanogaster*, Springer-Verlag, Berlin.

Carney et al. (1997) *Proc Natl Acad Sci USA* 94(22):12024-12029.

Casida & Quistad (1995) in Casida & Quistad, eds, Pyrethrum Flowers: Production, Chemistry, Toxicology, and Uses, pp.258-276, Oxford University Press, New York, New York.

Chibbar et al. (1993) *Plant Cell Rep* 12:506-509.

Christensen et al. (1989) *Plant Mol Biol* 12:619-632.

Christou et al. (1991) *Biotechnology* 9:957-962.

Coffer et al. (1996) *J Steroid Biochem Mol Biol* 58:467-477.

Conner et al. (1983) *Proc Natl Acad Sci USA* 80:278.

Cooper et al. (1996) *J Steroid Biochem Mol Biol* 57:251-257.

Cubitt et al. (1995) *Trends Biochem Sci* 20:448-455.

Dallery et al. (1993) *Biochem* 32:12428-12436.

Danielsen et al. (1989) *Cell* 57:1131-1138.

Datta et al. (1990) *Biotechnology* 8:736-740.

de Framond (1991) *FEBS* 290:103-106.

Demerec (1965) Biology of *Drosophila*, Hafner Publishing, New York, New York.

Desjarlais & Berg (1993) *Proc Natl Acad Sci USA* 90:2252-1860.

Devereux et al. (1984) *Nuc Acids Res* 12:387-395.

Dhadialla et al. (1998) *Annu Rev Entom* 43: 545-569.

Dittrich et al. (1997) *Development* 124(13):2515-2525.

Edwards & Kiehart (1996) *Development* 122:1499.

Eul et al. (1989) *EMBO J* 8:83-90.

European Patent Application No. 0 292 435

European Patent Application No. 0 332 104

European Patent Application No. 0 332 581

European Patent Application No. 0 342 926

European Patent Application No. 0 392 225

Evans (1988) *Science* 240:889-895.

Fire et al. (1998) *Nature* 391(6669):806-811.

Firek et al. (1993) *Plant Mol Biol* 22:129-142.

Forman & Samuels (1990) *Mol Endocrinol* 4:1293-1301.

Forman et al. (1989) *Mol Endocrinol* 3:1610-1626.

Freeman et al. (1999) *Development* 126:4591-4602.

Fromm et al. (1990) *Biotechnology* 8:833-839/

Giguere et al. (1987) *Nature* 330:624-629.

Glover, ed. (1985) DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, United Kingdom.

Goff et al. (1991) *Genes and Dev* 5:298-309.

Goldstein & Fyrberg, eds (1994) in Methods in Cell Biology, Vol. 44, Academic Press, San Diego, California.

Gordon-Kamm et al. (1990) *Plant Cell* 2:603-618.

Green & Chambon (1987) *Nature* 325:75-78.

Green et al. (1988) *EMBO J* 7:3037-3044.

Guichet et al. (1997) *Nature* 385:548-552.

Hadzic et al. (1995) *Mol Cell Biol* 15:4507-4517.

Hall et al. (2000) *Mol Enodcrinol* 14(12):2010-2023.

Hard et al. (1990) *Science* 249:397-404.

Harding & Lazar (1995) *Mol Cell Biol* 15(11):6479.

Harlow & Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Henikoff & Henikoff (1989) *Proc Natl Acad Sci USA* 89:10915.

Henikoff & Henikoff (2000) *Adv Protein Chem* 54:73-97.

Henikoff et al. (2000) *Electrophoresis* 21(9):1700-1706.

Henrich et al. (1987) *Dev Biol* 120(1):50-55.

Henrich et al. (1993) *Dev Genet* 14(5):469-477.

Higashijima et al. (1996) *Development* 122(2):527-536.

Hodgson & Tate (1976) in Wilkinson, ed, Insecticide Biochemistry and Physiology, pp. 115-148, Plenum Press, New York, New York.

Huang et al. (2000) *Pac Symp Biocomput* 230-241.

Hudspeth & Grula (1989) *Plant Mol Biol* 12:579-589.

Hunter (2000) *Curr Biol* 10(4):R137-140.

Hutchens & Yip (1993) *Rapid Commun Mass Spectrom* 7:576-580.

International Publication No. WO 00/37077

International Publication No. WO 93/07278

International Publication No. WO 93/21335

International Publication No. WO 94/00977

International Publication No. WO 97/47763

International Publication No. WO 99/26966

Johnson & Garza (1998) *Ann Dros Res Conf* 39:430A.

Johnston et al. (1997) *Mol Endocrinol* 11(3):342-352.

Karlin & Altschul (1993) *Proc Natl Acad Sci USA* 90:5873-5887.

Kennerdell & Carthew (2000) *Nat Biotech* 18(8):896-898.

Kestila et al. (1998) *Mol Cell* 1(4):575-582.

Kitts (1990) *Nucleic Acid Res* 18:5667.

Klein et al. (1987) *Nature* 327:70-73.

Koelle (1992) Ph.D. Thesis, Stanford University, Stanford, California.

Koelle et al. (1991) *Cell* 67:59-77.

Koziel et al. (1993) *Biotechnology* 11:194-200.

Kuwabara & Olson (2000) *Parasitol Today* 16(8):347-349.

Kyte et al. (1982) *J Mol Biol* 157:105.

Lam & Thummel (2000) *Curr Biol* 10(16):957-963.

Lam et al. (1999) *Dev Biol* 212(1):204-216.

Landegren et al. (1998) *Genome Res* 8:769-776.

Landgren et al. (1988) *Science* 241:1007.

Landgren et al. (1988) *Science* 242:229-237.

Lanz & Rusconi (1994) *Endocrinology* 135:2183-2194.

Laudet et al. (1992) *EMBO J* 11:1003-1013.

Lazar & Harding (1998) in Freedman, ed, Molecular Biology of Steroid and Nuclear Hormone Receptors, pp. 269-270, Birkhäuser, Boston, Massachusetts.

Lebel et al. (1998) *Plant J* 16:223-233.

Leng et al. (1995) *Mol Cell Biol* 15:255-263.

Liedberg et al. (1983) *Sensors Actuators* 4:299-304.

Lin et al. (1991) *Mol Endocrinol* 5:485-492.

Lindsley & Zimm (1992) The Genome of *Drosophila melanogaster*, Academic Press, San Diego, California.

Lis et al. (1983) *Cell* 35:403.

Liu et al. (1997) *Proc Natl Acad Sci USA* 94:5525-5530.

Logemann et al. (1989) *Plant Cell* 1:151-158.

Lowman (1997) *Annu Rev Biophys Biomol Struct* 26:401-424.

Lucknow & Summers (1988) *Bio/Technol* 6:47.

Lupisella et al. (1995) *J Biol Chem* 270:24884-24890.

Maalouf et al. (1998) *J Biomol Struc Dynamics* 15(5):841-851.

Mader et al. (1993) *EMBO J* 12:5029-5041.

Madge et al. (1972) *Phys Rev Lett* 29:705-708.

Maiti et al. (1997) *Proc Natl Acad Sci USA* 94:11753-11757.

Mak et al. (1989) *J Biol Chem* 264:21613:21618.

Malmquist (1993) *Nature* 361:186-187.

McEclroy et al. (1991) *Mol Gen Genet* 231:150-160.

McElroy et al (1990) *Plant Cell* 2:163-171.

Meijer (2000) *Mol Gen Genet* 263:12-21.

Meshi (1995) *Plant Cell Physiol* 36:1405-1420.

Messing & Vierra (1982) *Gene* 19: 259-268.

Misquitta & Paterson (1999) *Proc Natl Acad Sci USA* 96 (4):1451-1456.

Nauber et al. (1988) *Nature* 336(6198):489-492.

Needleman & Wunsch (1970) *J Mol Biol* 48:443.

Nemoto et al. (1992) *J Steroid Biochem Mol Biol* 42:803-812.

Ng (2000) *Trends Biochem Sci* 25:121-126.

Nijhout (1994) Insect Hormones, Princeton University Press, Princeton, New Jersey.

Norris et al. (1993) *Plant Mol Biol* 21:895-906.

Northrop et al. (2000) *Mol Endocrinol* 14(5)605-622.

Ohtsuka et al. (1985) *J Biol Chem* 260:2605-2608.

Orita et al. (1989) *Proc Natl Acad Sci USA* 86(8):2766-2770.

Oro et al. (1988) *Nature* 336(61989):493-496.

Oro et al. (1992) *Development* 115(2):449-462.

Pakdel et al. (1993) *Mol Endocrinol* 7:1408-1417.

Paszkowski et al. (1984) *EMBO J* 3:2717-2722.

Pearson & Lipman (1988) *Proc Natl Acad Sci USA* 85:2444-2448.

Perlmann et al. (1993) *Genes Dev* 7:1411-1422.

Perrimon et al. (1985) *Genetics* 11:23-41.

Perry et al. (1997) Insecticides in Agriculture and Environment: Retrospects and Prospects, pp. 7-10, Springer-Verlag, New York, New York.

Perry et al. (1997) Insecticides in Agriculture and Environment: Retrospects and Prospects, pp. 12-13, Springer-Verlag, New York, New York.

Potrykus et al. (1985) *Mol Gen Genet* 199:169-177.

Ptashne (1988) *Nature* 335:683-689.

Rastinejad (1998) in Freedman, ed, Molecular Biology of Steroid and Nuclear Hormone Receptors, pp. 107, Birkhäuser, Boston, Massachusetts.

Rastinejad (1998) in Freedman, ed, Molecular Biology of Steroid and Nuclear Hormone Receptors, pp. 103-131, Birkhäuser, Boston, Massachusetts.

Rastinejad et al. (1995) *Nature* 375:203-211.

Reich et al. (1986) *Biotechnology* 4:1001-1004.

Roberts (1986) *Drosophila melanogaster*, A Practical Approach, IRL Press, Washington, DC.

Rochel et al. (1997) *Biochem Biophys Res Commun* 230:293-296.

Rohrmeier & Lehle (1993) *Plant Mol Biol* 22:783-792.

Rose and Botstein (1983) *Meth Enzymol* 101:167-180.

Rossolini et al. (1994) *Mol Cell Probes* 8:91-98.

Rusconi & Yamamoto (1987) *EMBO J* 6:1309-1315.

Sadowski et al. (1988) *Nature* 335:563-564.

Saiki et al. (1985) *Bio/Technology* 3:1008-1012.

Sambrook et al., eds (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Saqi et al. (1999) *Bioinformatics* 15(6):521-522.

Schapira et al. (2000) *Proc Natl Acad Sci USA* 97(3):1008-1013.

Schneuwly et al. (1987) *Nature* 325:816.

Segraves & Hogness (1990) *Genes Dev* 4:204-219.

Sharp (1999) Genes Dev 13(2):139-141.

Shimamoto et al. (1989) *Nature* 338:274-277.

Sidhu (2000) *Curr Opin Biotech* 11(6):610-616.

Silhavy et al. (1984) Experiments with Gene Fusions, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Simental et al. (1991) *J Biol Chem* 266:510-518.

Simons in Freedman, ed, Molecular Biology of Steroid and Nuclear Hormone Receptors, pp. 35-104, Birkhäuser, Boston, Massachusetts.

Sliter & Gilbert (1992) *Genetics* 130:555-568.

Smith & Waterman (1981) *Adv Appl Math* 2:482.

Spencer et al. (1990) *Theor Appl Genet* 79:625-631.

Stoneking et al. (1991) *Am J Hum Genet* 48(2):370-82.

Struhl (1985) *Nature* 318:677.

Summers & Smith *Texas Agriculture Exp Station Bulletin* No. 1555.

Sutherland et al. (1995) *Proc Natl Acad Sci USA* 92(17):7966-7970.

Taylor et al (1993) *Plant Cell Rep* 12:491-495.

Tetel et al. (1997) *Mol Endocrinol* 11:1114-1128.

Thomas et al. (1993) *Nature* 362:471-475.

Thompson et al. (1989) *Proc Natl Acad Sci USA* 86:3493-3498.

Thummel (1995) *Cell* 83:1-20.

Thummel (1997) *BioEssays* 19(8):669-672.

Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, part I chapter 2, Elsevier, New York, New York.

Triezenberg et al. (1988) *Genes Dev* 2(6):718-729.

Tsai et al. (1998) in Wilson et al., eds, Williams Textbook of Endocrinology, pp. 55-94, W.B. Saunders Company, Philadelphia, Pennsylvania.

U.S. Patent No 4,196,265

U.S. Patent No. 4,554,101

U.S. Patent No. 4,940,935

U.S. Patent No. 4,945,050

U.S. Patent No. 4,981,784

U.S. Patent No. 5,036,006

U.S. Patent No. 5,100,792

U.S. Patent No. 5,188,642

U.S. Patent No. 5,260,203.

U.S. Patent No. 5,466,785

U.S. Patent No. 5,510,240

U.S. Patent No. 5,591,616

U.S. Patent No. 5,686,600

U.S. Patent No. 5,767,378

U.S. Patent No. 5,834,228

U.S. Patent No. 5,872,011

U.S. Patent No. 5,880,333

U.S. Patent No. 5,994,629

Ueda et al. (1992) *Mol Cell Biol* 12(12):5667-5672.

Uknes et al. (1992) *Plant Cell* 4:645-656.

Umesono & Evans (1989) *Cell* 57:1139-1146.

Vasil et al. (1992) *Biotechnology* 10:667-674.

Vasil et al. (1993) *Biotechnology* 11:1553-1558.

Wagner et al. (1995) *Nature* 378:690-697.

Wang et al. (1998) *Science* 280(5366):1077-1082.

Warner et al. (1993) *Plant J* 3:191-201.

Weeks et al. (1993) *Plant Physiol* 102:1077-1084

Wentworth et al. (2000) *J Endocrinol* 166(3):R11-16.

White et al. (1990) *Nuc Acids Res* 18:1062.

White et al. (1997) *Science* 276(5309):114-117.

Wilkinson (1976a) in Tahori, ed, Proc 2nd Int Congr on Pesticides and Chemistry, Vol. 2, pp. 117-159, Gordon & Breach, New York, New York.

Wilkinson (1976b) in Metcalf & McKelvey, eds, The Future for Insecticides: Needs and Prospects, Vol. 6, pp. 191-178, Wiley, New York, New York.

Wilson et al. (1992) *Science* 256:106-110.

Wilson et al. (1993) *Mol Cell Biol* 13(9):5794-5804.

Wing (1988) *Science* 241:467-469.

Woodard et al. (1994) *Cell* 79(4):607-615.

Worrall et al. (1998) *Anal Biochem* 70:750-756.

Wu (2000) *Plant J* 22:19-27.

Wurtz et al. (1996) *Nat Struct Biol* 3:87-94.

Xu et al. (1993) *Plant Mol Biol* 22:573-588.

Xu et al. (1996) *J Biol Chem* 271:21430-21438.

Yamamoto (1973) in Casida, ed, Pyrethrum, The Natural Insecticide, pp. 191-170, Academic Press, New York, New York.

Yang & Chen (1999) *Cancer Res* 59(18):4519-4524.

Yao (1993) *Cell* 71:63-72.

Yu et al. (1997) *Nature* 385:552-555.

Yuan et al. (1999) *Hum Mutat* 14(5):440-446.

Zechel et al. (1994) *EMBO J* 13:1414-1424.

Zhang et al. (1988) *Plant Cell Rep* 7:379-384.

Zhang et al. (1996a) *Mol Endocrinol* 10:24-34.

Zhang et al. (1996b) *J Biol Chem* 271:14825-14833.

Zimmer et al. (1993) Peptides, pp. 393–394, ESCOM Science Publishers, B. V.

It will be understood that various details of the invention can be changed without departing from the scope of the invention. Furthermore, the foregoing description is for the purpose of illustration only, and not for the purpose of limitation--the invention being defined by the claims appended hereto.

CLAIMS

What is claimed is:

1. An isolated insect nuclear receptor polypeptide comprising:

(a) a polypeptide encoded by the nucleotide sequence of any one of SEQ ID NOs:1, 5, 9, 17, 19, 21, 23, and 25;

(b) a polypeptide encoded by a nucleic acid molecule that is substantially identical to any one of SEQ ID NOs:1, 5, 9, 17, 19, 21, 23, and 25;

(c) a polypeptide comprising the amino acid sequence of any one of SEQ ID NOs:2, 6, 10, 18, 20, 22, 24, and 26;

(d) a polypeptide that is a biological equivalent of the polypeptide of any one of SEQ ID NOs:2, 6, 10, 18, 20, 22, 24, and 26; or

(e) a polypeptide which is immunologically cross-reactive with an antibody that shows specific binding with a polypeptide of any one of SEQ ID NOs:2, 6, 10, 18, 20, 22, 24, and 26.

2. An isolated nucleic acid molecule encoding an insect nuclear receptor polypeptide comprising:

(a) a nucleotide sequence of any one of SEQ ID NOs:1, 5, 9, 17, 19, 21, 23, and 25; or

(b) a nucleic acid molecule substantially identical to any one of SEQ ID NOs:1, 5, 9, 17, 19, 21, 23, and 25.

3. A chimeric gene, comprising the nucleic acid molecule of claim 2 operatively linked to a heterologous promoter.

4. A vector comprising the chimeric gene of claim 3.

5. A host cell comprising the chimeric gene of claim 3.

6. The host cell of claim 5, wherein the cell is selected from the group consisting of a bacterial cell, an insect cell, and a plant cell.

7. A method of detecting a nucleic acid molecule that encodes an insect nuclear receptor polypeptide, the method comprising:

(a) procuring a biological sample comprising nucleic acid material;

(b) hybridizing the nucleic acid molecule of claim 2 under stringent hybridization conditions to the biological sample of (a), thereby forming a duplex structure between the nucleic acid of claim 2 and a nucleic acid within the biological sample; and

(c) detecting the duplex structure of (b), whereby a nuclear receptor nucleic acid molecule is detected.

8. An antibody that specifically recognizes an insect nuclear receptor polypeptide of claim 1.

9. A method for producing an antibody that specifically recognizes an insect nuclear receptor polypeptide, the method comprising:

(a) recombinantly or synthetically producing an insect nuclear receptor polypeptide, or portion thereof, as set forth in any of SEQ ID NOs:2, 6, 10, 18, 20, 22, 24, and 26;

(b) formulating the polypeptide of (a) whereby it is an effective immunogen;

(c) administering to an animal the formulation of (b) to generate an immune response in the animal comprising production of antibodies, wherein antibodies are present in the blood serum of the animal; and

(d) collecting the blood serum from the animal of (c), the blood serum comprising antibodies that specifically recognize a nuclear receptor polypeptide.

10. A method for detecting a level of nuclear receptor polypeptide, the method comprising

(a) obtaining a biological sample comprising peptidic material; and

(b) detecting a nuclear receptor polypeptide in the biological sample of (a) by immunochemical reaction with the antibody of claim 8, whereby a level of nuclear receptor polypeptide in a sample is determined.

11. A method for identifying a substance that modulates nuclear receptor function, the method comprising:

(a) isolating an insect nuclear receptor polypeptide of claim 1;

(b) exposing the isolated insect nuclear receptor polypeptide to a plurality of candidate substances;

(c) assaying binding of a candidate substance to the isolated nuclear receptor polypeptide; and

(d) selecting a candidate substance that demonstrates specific binding to the isolated insect nuclear receptor polypeptide.

12. A method for identifying an insecticidal substance that modulates nuclear receptor function, the method comprising:

(a) isolating an insect nuclear receptor polypeptide of any one of even numbered SEQ ID NOs:2-34, wherein modulation of the insect nuclear receptor polypeptide confers lethality of an insect during a larval stage;

(b) exposing the isolated insect nuclear receptor polypeptide to a plurality of substances;

(c) assaying binding of a substance to the isolated nuclear receptor polypeptide; and

(d) selecting a substance that demonstrates specific binding to the isolated insect nuclear receptor polypeptide.

13. A method for preventing or abrogating an insect infestation of a plant, the method comprising:

(a) preparing an insecticidal composition that includes an insect nuclear receptor modulator identified according to the method of claim 12; and

(b) contacting an effective dose of the insecticidal composition with a plant, whereby an insect infestation of a plant is prevented or abrogated.

14. The method of claim 13, wherein the insecticidal composition comprises a chemical compound, a protein, a peptide, a nucleic acid, or an antibody.

15. A method for preventing or abrogating a nematode infestation of a plant, the method comprising:

(a) preparing an insecticidal composition that includes an insect nuclear receptor modulator identified according to the method of claim 12; and

(b) contacting an effective dose of the insecticidal composition with a plant, whereby a nematode infestation of a plant is prevented or abrogated.

16. A method for preventing or abrogating an insect infestation of a plant, the method comprising expressing in a plant an insect nuclear receptor modulator that modulates the activity of an insect nuclear receptor polypeptide of claim 1, whereby an insect infestation of the plant is prevented or abrogated.

17. The method of claim 16, wherein the bioactive agent comprises a protein, a peptide, a nucleic acid, or an antibody.

18. A method for preventing or abrogating a nematode infestation of a plant, the method comprising expressing in a plant a bioactive agent that modulates the activity of an insect nuclear receptor polypeptide of claim 1, whereby a nematode infestation of the plant is prevented or abrogated.

19. A chimeric nuclear receptor cassette comprising a DNA binding domain, a ligand binding domain, and an activation or repression domain, wherein one or more of the DNA binding domain, the ligand binding domain, and the activation domain comprises an amino acid sequence that is identical or substantially identical to a portion of any one of SEQ ID NOs:2, 6, 10, 18, 20, 22, 24, and 26.

20. A method of inducing expression of a target nucleotide sequence, the method comprising:

(a) constructing a chimeric nuclear receptor expression cassette of claim 19; and

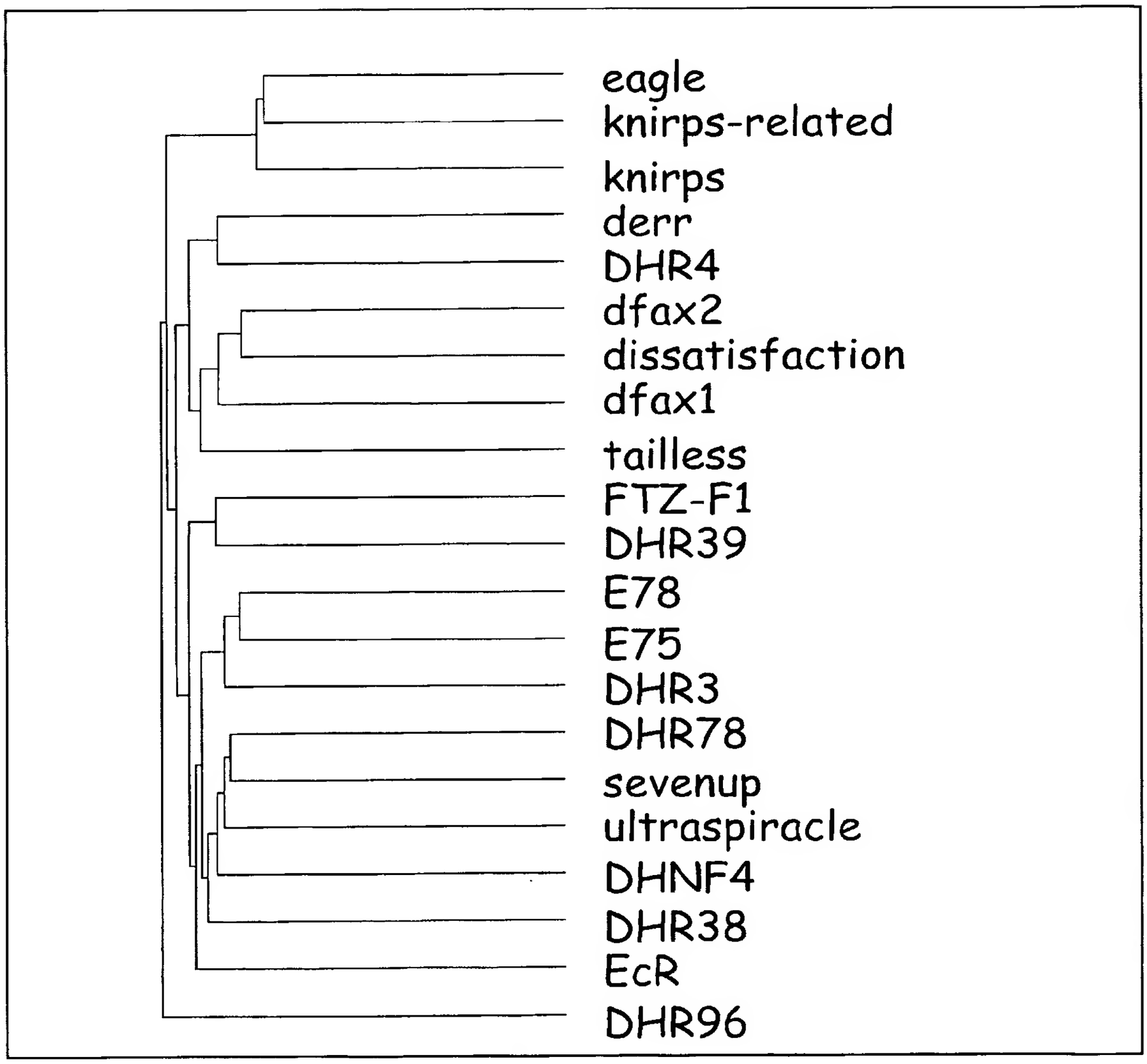
(b) constructing a target expression cassette having a target nucleotide sequences and a cis-regulatory element that is recognized by a DNA-binding domain of the chimeric nuclear receptor expression cassette;

(c) expressing the chimeric nuclear receptor expression cassette and the target expression cassette in a heterologous organism; and

(d) contacting a ligand that binds to the ligand binding domain of the chimeric nuclear receptor expression cassette with the organism, whereby the target nucleotide sequence is expressed.

21. The method of claim 20, wherein the heterologous organism is a plant.

FIG. 1


eagle
knirps-related
knirps
derr
DHR4
dfax2
dissatisfaction
dfax1
tailless
FTZ-F1
DHR39
E78
E75
DHR3
DHR78
sevenup
ultraspiracle
DHNF4
DHR38
EcR
DHR96

```
B.mori          ----------------------------------------------------MSSVAK
M.sexta         ----------------------------------------------------MSSVAK
C.fumiferana    ----------------------------------------------------MSSVAK
H.virescens     ----------------------------------------------------------
L.migratoria    ----------------------------------------------------------
C.tentans       MLKKEKPMMTVAAIIEQAQNRWMDHPLVYNSRSLQFQGSYCIDSSLLGHMGPLSPPDL
D.melanogaster  ----------------------------------------------------------

B.mori          KDKRTMSVTALINRAWPMTPSPQQQQQMVPSTQ...HSNFLHAMATPSTTPNVE..LD
M.sexta         KDKRTMSVTALINRAWPLTPAPHQQQSM.PSSQ...PSNFLQPLATPSTTPSVE..LD
C.fumiferana    KDKPTMSVTALINWARPAPPGPPQPQSASPAPAAMLQQLPTQSMQSLNHIPTVDCSLD
H.virescens     ----------------------------------------------------MMEPSRD
L.migratoria    ----------------------------------------------------MEGSERGIS
C.tentans       KPDISLLNCNNNNNNTNNNNSNSSHNNLNHHNTSPLPVLGANTFSPIQSLNNNGPSSP
D.melanogaster  ----------------------------------MDNCDQDASFRLSHIKEEVKPD

B.mori          IQWLN.IESGFMSPMSPPEMK.PDTAMLDGFRDDS.....TPPPPF..........
M.sexta         IQWLN.IEPGFMSPMSPPEMK.PDTAMLDGLRDDS.....TPPPAF..........
C.fumiferana    MQWLN.LEPGFMSPMSPPEMK.PDTAMLDGLRDDA.....TSPPNF..........
H.virescens     .SGLN.LEGGFMSPMSPPEMK.PDTAMLDGLRDDS.....TPPPAF..........
L.migratoria    LE.NN.LSISSMGPQSPLDMK.PDTASLISSGSFSPTGGPNSPGSFTIGHSSLLNNSS
C.tentans       LSSIG.NGSGTIVTFNQIKLQSPSPSNASSSSTLSGPLTTTPPATNANNILGMGNGNC
D.melanogaster  ISQLNDSNNSSFSPKAESPVPFMQAMSMVHVLPGSNSASSNNNSAGDAQMAQAPNSAG

B.mori          ........KNYPPNHPLSGSKHLCSICGDRASGKHYGVYSCEGCKGFFKRTVRKDLTY
M.sexta         ........KNYPPNHPLSGSKHLCSICGDRASGKHYGVYSCEGCKGFFKRTVRKDLTY
C.fumiferana    ........KNYPPNHPLSGSKHLCSICGDRASGKHYGVYSCEGCKGFFKRTVRKDLSY
H.virescens     ........KNYPPNHPLSGSKHLCSICGDRASGKHYGVYSCEGCKGFFKRTVRKDLTY
L.migratoria    SNQAKGSSSQYPPNHPLSGSKHLCSICGDRASGKHYGVYSCEGCKGFFKRTVRKDLSY
C.tentans       GNTANGKQSQYPPNHPLSGSKHLCSICGDRASGKHYGVYSCEGCKGFFKRTVRKDLSY
D.melanogaster  GSAAAAVQQQYPPNHPLSGSKHLCSICGDRASGKHYGVYSCEGCKGFFKRTVRKDLTY

B.mori          ACREDKNCIIDKRQRNRCQYCRYQKCLACGMKREAVQEERQRAARRTE.DAHPS....
M.sexta         ACREDRNCIIDKRQRNRCQYCRYQKCLACGMKREAVQEERQRAARGTE.DAHPS....
C.fumiferana    ACREERNCIIDKRQRNRCQYCRYQKCLACGMKREAVQEERQRNARGAE.DAHPS....
H.virescens     ACREERNCIIDKRQRNRCQYCRYQKCLACGMKREAVQEERQRAARGTE.DAHPS....
L.migratoria    ACREDKNCIIDKRQRNRCQYCRYQKCLAMGMKREAVQEERQRTKERDQ.NEVES....
C.tentans       ACREERNCVIDKKQRNRCQYCRYQKCLNCGMKREAVQEERQRGGKSQKGDDMSI.....
D.melanogaster  ACRENRNCIIDKRQRNRCQYCRYQKCLTCGMKREAVQEERQRGARNAAGRLSASGGGS

B.mori          ........SSVQ.............................ELSIERLLELEAL..VAD
M.sexta         ........SSVQ.............................ELSIERLLEIESL..VAD
C.fumiferana    ........SSVQ.....................V......SDELSIERLTEMESL..VAD
H.virescens     ........SSVQ.....................V.......QELSIERLLEMESL..VAD
L.migratoria    ........TS.S.....................L......HTDMPVERILEAEKR..VE.
C.tentans       ........SSTQ.....................SLVNNGPGRDITVERLMEADQMSEARC
D.melanogaster  SGPGSVGGSSSQGGGGGGGVSGGMGSGNGSDDFMTNSVSRDFSIERIIEAEQRAETQC

B.mori          SAEELQILRVGPESG.VPAKYRAPVSSLCQIGNKQIAALIVWARDIPHFGQLEIDDQI
M.sexta         PPEEFQFLRVGPESG.VPAKYRAPVSSLCQIGNKQIAALVVWARDIPHFGQLELEDQI
C.fumiferana    PSEEFQFLRVGPDSN.VPPRYRAPVSSLCQIGNKQIAALVVWARDIPHFGQLELDDQV
H.virescens     PSEEFQFLRVGPDSN.VPPKFRAPVSSLCQIGNKQIAALVVWARDIPHFSQLEMEDQI
L.migratoria    ..........................CKAENQVEYELVEWAKHIPHFTSLPLEDQV
C.tentans       GDKSIQYLRVAASNTMIPPEYRAPVSAICAMVNKQVFQHMDFCRRLPHFTKLPLNDQM
D.melanogaster  GDRALTFLRVGPYST.VQPDYKGAVSALCQVVNKQLFQMVEYARMMPHFAQVPLDDQV

B.mori          LLIKGSWNELLLFAIAWRSMEFLNDERENV.......D...SRNT..APPQLICLMPG
M.sexta         LLIKNSWNELLLFAIAWRSMEYLTDERENV.......D...SRST..APPQLMCLMPG
C.fumiferana    VLIKASWNELLLFAIAWRSMEYLEDERENG.......DGT..RST..TQPQLMCLMPG
```

Figure 2A

```
H.virescens       LLIKGSWNELLLFAIAWRSMEFLTEERDGV.....DGTGNRTT..SPPQLMCLMPG

L.migratoria      LLLRAGWNELLIAAFSHRSV....DVKDG.....................IVLATG
C.tentans         YLLKQSLNELLILNIAYMSIQYVEPDRRNA.....DGSLERRQ..ISQQ.MCLSRN
D.melanogaster    ILLKAAWIELLIANVAWCSIVSLDDGGAGGGGGGLGHDGSFERRSPGLQPQQLFLNQS

B.mori            MTLHRNSALQAGVGQIFDRVLSELSLKMRSLRMDQAECVALKAIILLNPDVKGLKNKQ
M.sexta           MTLHRNSALQAGVGQIFDRVLSELSLKMRTLRMDQAEYVALKAIILLNPDVKGLKNKP
C.fumiferana      MTLHRNSAQQAGVGAIFDRVLSELSLKMRTLRMDQAEYVALKAIVLLNPDVKGLKNRQ
H.virescens       MTLHRNSALQAGVGQIFDRVLSELSLKMRTLRVDQAEYVALKAIILLNPDVKGLKNRQ
L.migratoria      LTVHRNSAHQAGVGTIFDRVLTELVAKMREMKMDKTELGCLRSVILFNPEVRGLKSAQ
C.tentans         YTLGRNMAVQAGVVQIFDRILSELSVKMKRLDLDATELCLLKSIVVFNPDVRTLDDRK
D.melanogaster    FSYHRNSAIKAGVSAIFDRILSELSVKMKRLNLDRRELSCLKAIILYNPDIRGIKSRA

B.mori            EVDVLREKMFLCLDEYCRRSRGGEEGRFAALLLRLPALRSISLKSFEHLYLFHLVAEG
M.sexta           EVVVLREKMFSCLDEYVRRSRCAEEGRFAALLLRLPALRSISLKCFEHLYFFHLVADT
C.fumiferana      EVDVLREKMFSCLDDYCRRSRSNEEGRFASLLLRLPALRSISLKSFEHLYFFHLVAEG
H.virescens       EVEVLREKMFLCLDEYCRRSRSSEEGRFAALLLRLPALRSISLKSFEHLFFFHLVADT
L.migratoria      EVELLREKVYAALEEYTRTTHPDEPGRFAKLLLRLPSLRSIGLKCLEHLFFFRLIGDV
C.tentans         SIDLLRSRIYASLDEYCRQKHPNEDGRFAQLLLRLPALRSISLKCLDHLFYFQLIDDK
D.melanogaster    EIEMCREKVYACLDEHCRLEHPGDDGRFAQLLLRLPALRSISLKCQDHLFLFRITSDR

B.mori            SVSSYIRDALCNHAPPIDTNIM----------------------------------
M.sexta           SIASYIHDALRNHAPSIDTSIL----------------------------------
C.fumiferana      SISGYIREALRNHAPPIDVNAMM---------------------------------
H.virescens       SIAGYIRDALRNHAPPIDTNMM----------------------------------
L.migratoria      PIDTFLMEMLESPSDS----------------------------------------
C.tentans         NVENSVIEEFHKLN------------------------------------------
D.melanogaster    PLEELFLEQLEAPPPPGLAMKLE---------------------------------
```

Figure 2B

```
hv.bftz   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
bmftz     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
dmftz     MDTFNVPMLAESSNTNYATEATSNHHHLQHQHQQQHSHQQQQQQQQQLLMPHHHKDQMLA

hv.bftz   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
bmftz     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
dmftz     AGSSPMLPFYSHLQLQQKDATATIGPAAAAAAVEAATTSANADNFSSLQTIDASQLDGGI

hv.bftz   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
bmftz     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
dmftz     SLSGLCDRFFVASPNPHSNSNMTLMGTATAATTTTTNNNNNNNTNNNNNNNVEAKTVRPS

hv.bftz   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
bmftz     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
dmftz     NGNSVIIESVTMPSFANILFPTHRSANECIDPALLQKNPQNPNGNNSSIIVPPVEYHQLK

hv.bftz   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
bmftz     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
dmftz     PLEVNSSTSVSTSNFLSSTTAQLLDFEVQVGKDDGHISTTTTTGPGSGSASGSGSGSGSG

hv.bftz   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
bmftz     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
dmftz     SGSIASTIGTATPTTTTSMSNTANPTRSSLHSIEELAASSCAPRAASPNSNHTSSASTTP

hv.bftz   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MTMDQQTGLMSLNMSPFDLSPGPEGSG
bmftz     ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MHEDAPKMSIAQSLAAS
dmftz     QQQQQQQHHMQSGNHSGSNLSSDDESMSEDEFGLEIDDNGGYQDTTSSHSQQSGGGGGGG

hv.bftz   SGGGPSSASQQYVPQGAAYQCPPEQQSFGYAN..LDASYLFP.T...........GAGGE
bmftz     TSQPKGDIVTEIPLEFA.MSSMETKSIETTNVELKITYVDPTT...........GTGGEP
dmftz     GGNLLNGSSGGSSAGGGYMLLPQAASSSGNNGNPNAGHMSSGSVGNGSGGAGNGGAGGNS

hv.bftz   PGAYLPAAGTVC......DQTDTKDVIEELCPVCGDKVSGYHYGLLTCESCKGFFKRTVQ
bmftz     GAYLPTAGTVC......DQTDTKDVIEELCPVCGDKVSGYHYGLLTCESCKGFFKRTVQN
dmftz     GPGNPMGGTSATPGHGGEVIDFKHLFEELCPVCGDKVSGYHYGLLTCESCKGFFKRTVQN

hv.bftz   NKKVYTCVAERACHIDKTQRKRCPFCRFQKCLDVGMKLEAVRADRMRGGRNKFGPMYKRD
bmftz     KKVYTCVAERACHIDKTQRKRCPFCRFQKCLDVGMKLEAVRADRMRGGRNKFGPMYKRDR
dmftz     KKVYTCVAERSCHIDKTQRKRCPYCRFQKCLEVGMKLEAVRADRMRGGRNKFGPMYKRDR

hv.bftz   RARKLQMMRQRQIAVQTLRGSLG....DSGLVLGFGSPYATVPVKQEIQIPQVSSLTSSP
bmftz     ARKLQMMRQRQIAVQTLRGSLG....DGGLVLGFGSPYTAVSVKQEIQIPQVSSLTSSPE
dmftz     ARKLQVMRQRQLALQALRNSMGPDIKPTPISPGYQQAYPNMNIKQEIQIPQVSSLTQSPD

hv.bftz   ESSPGPALLA............AQPQPP..........QPPP..................
bmftz     SSPGPALLR............AQPQPP..........QPPP...................
dmftz     SSPSPIAIALGQVNASTGGVIATPMNAGTGGSGGGGLNGPSSVGNGNSSNGSSNGNNNSS

hv.bftz   .....................PPAHDKW......EAHSPHSASPDAFAFDAPA....TAA
bmftz     ....................PPTHDKW......EAHSPHSASPDAFTFDTQS....NTAA
dmftz     TGNGTSGGGGGNNAGGGGGGTNSNDGLHRNGGNDSSSCHEAGIGSLQNTADSKLCFDSGT

hv.bftz   ATPSSTAEPTSTETLRVSPMIREFVQTIDDREWQNSLFGLLQSQTYNQCEVDLFE.LMCK
bmftz     TPSSTAEATSTETLRVSPMIREFVQTVDDREWQNALFGLLQSQTYNQCEVDLFE.LMCKV
dmftz     HPSSTADAL.IEPLRVSPMIREFVQSIDDREWQTQLFALLQKQTYNQVEVDLFELLMCKV

hv.bftz   VLDQNLFSQVDWARNTVFFKYLKVDDQMKLLQHSWSDMLVLDHLHQRMHNGLPDETTLHN
bmftz     LDQNLFSQVDWARNTVFFKYLKVDDQMKLLQDSWSVMLVLDHLHQRMHNGLPDETTLHNG
dmftz     LDQNLFSQVDWARNTVFFKDLKVDDQMKLLQHSWSDMLVLDHLHHRIHNGLPDETQLNNG
```

Figure 3A

```
hv.bftz    GQKFDLLCLGLLGVPSLADHFNELQNKLAELKFDVPDYICVKFLLLLNPEVRGIVNVKCV
bmftz      QKFDLLCLGLLGVPSLADHFNELQNKLAELKFDVPDYICVKFMLLLNPEVRGIVNVKCVR
dmftz      QVFNLMSLGLLGVPQPGDYFNELQNKLQDLKFDMGDYVCMKFLILLNPSVRGIVNRKTVS

hv.bftz    RDGYQTVQAALLDYTLSCYPTIQDKFDKLVMVVPEIHALAARGEEHLYQRHCAGQAPTQT
bmftz      EGYQTVQAALLDYPY.LLSTIQDKFGKLVMVVPEIHALRL.GEKST....CTSGIVQARH
dmftz      EGHDNVQAALLDYTLTCYPSVNDKFRGLVNILPEIHAMAVRGEDHLIT..CTPSTVPAVR

hv.bftz    LLMEMLHAKRKS-----------------------------------------------
bmftz      LPRLFSWKCCTQNANLEVPVTNKVEELRSAKPRRHHNK--------------------
dmftz      PPKRCSWRCCTPSARDRGRENVTRNT---------------------------------
```

Figure 3B

```
D.melanogaster(A)   ~~~~ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(A)          ~~~~~ ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~ ~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   MHGGGPGSSGSNIIRRSSGSFPGSGSGSASKLIKTEPIDFEMLHLEENERQQDIEREPSS
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~MVCAMQEVAAVQHQQQQQQLQLPQQQQQQQQTTQQQHATTIVLLTGNGGGNLHIVA
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   SNSNSNSNSLTPQRYTHVQVQTVPPRQPTGLTTPGGTQKVILTPRVEYVQQRATSSTGGG
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   TPQQHQPMHQLHHQHQHQHQHQQQAKSQQLKQQHSALVKLLESAPIKQQQQTPKQIVYLQ
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   MKHVYSQQQGTAASRSAPPETTALLTTTSGTPQIIITRTLPSNQHLSRRHSASPSALHHY
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MLMSADSSDSAKTSVICSTVSAS
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   QQQQQPQRKRLKNEAAIVQQQQQTPATLVKTTTTSNSNSNNTQTTNSISQQQQQHQIVLQ
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   QQQQQPQRQQSPPPLHHQQQQQQQHVRVIRDGRLYDEATVVVAARRHSVSPPPLHHHSRS
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   MLAPPAPEQPSTTAPPILGVTGRSHLENALKLPPNTSVSAYYQHNSKLGMGQNYNPEFRS
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   HQQPAAAATPKPCADLSAKNDSESGIDEDCPNSDEDCPNANPAGTSLEDSSYEQYQCPWK
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   APVSPVIARRGGAAAYMDQQYQQRQTPPLAPPPPPPPPPPPPPPPQQQQQQYISTGVPPP
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   LVAPVTDLDTVPPTGVTMASSSNSPNSSVKLPHSGVIFVSKSSAVSTTDGPTAVLQQQQP
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   KIRYARELLKQRELEQQQTTGGSNAQQQVEAKPAAIPTSNIKQLHCDSPFSAQTHKEIAN
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   TAAARKFVVSTSTRHVNVIASNHFQQQQQQHQAQQHQQQHQQHVIASVSSSSSSSAIGSG
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Figure 4A

```
D.melanogaster(A)   QQQMPQHFESLPHHHPQQEHQPQQQQQQHHLQHHPHPHVMYPHGYQQANLHHSGGIAVVP
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   LLRQQSQQQQVVATQQQQQQQQQHQHQQQRRDSSDSNCSLMSNSSNSSAGNCCTCNAGDD
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   GSSSSHIFRTPVVSSSSSSNMHHQQQQQQQQSSLGNSVMRPPPPPPPPKVKHAS......
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MTLVMSP
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~MTLVMSP
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~PPAARAPSARLMRLPPLPLPDMTVTECQRRLLEP

D.melanogaster(A)   ADSRPQTPEYIKSYPVMDTTVASSVKGEPELNIEFDGTTVLCRVCGDKASGFHYGVHSCE
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~FDGTTVLCRVCGDKASGFHYGVHSCE
D.melanogaster(B)   QQLEEMDEAHDSGCDDELCEQHHQRLDSSQLNYLCQKFDEKLDTALSNSSA.NTG.RNTP
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~MVRAMSCGAELRERHSVLVSMLEARRESSDSGCSSDDGSDVER
D.melanogaster(C)   SSSSGNSSSSNTNNSSSSSNGEEPSSSIPDL..EFDGTTVLCRVCGDKASGFHYGVHSCE
C.fumiferana        DSSYGRYDAQPPVDGGMVNPVHR..EREPELHIEFDGTTVLCRVCGDKASGFHYGVHSCE
G.mellonella        DSSYGRYDAPAPADNRIMSPVHK..EREPELHIEFDGTTVLCRVCGDKASGFHYGVHSCE
M.ensis             ~~~~~~~~~~~MFCDQDMYEIPADCQVLVDKTVIEFDGTTVLCRVCGDKASGFHYGVHSCE
H.virescens         SAAEPPPPAPPTDSDVLLGRV.........LAEFDGTTVLCRVCGDKASGFHYGVHSCE

D.melanogaster(A)   GCK.....GFFRRSIQQKIQYRPCTKNQQCSILRINRNRCQYCRLKKCIAVGMSRDAVR
M.sexta(A)          GCK~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   AVTANEDADGFFRRSIQQKIQYRPCTKNQQCSILRINRNRCQYCRLKKCIAVGMSRDAVR
M.sexta(B)          DCKCRCDPQGFFRRSIQQKIQYRPCTKNQQCSILRINRNRCQYCRLKKCIAVGMSRDAVR
D.melanogaster(C)   GCK.....GFFRRSIQQKIQYRPCTKNQQCSILRINRNRCQYCRLKKCIAVGMSRDAVR
C.fumiferana        GCK.....GFFRRSIQQKIQYRPCTKNQQCSILRINRNRCQYCRLKKCIAVGMSRDAVR
G.mellonella        GCK.....GFFRRSIQQKIQYRPCTKNQQCSILRINRNRCQYCRLKKCIAVGMSRDAVR
M.ensis             GCK.....GFFRRSIQQKIQYRPCTKNQQCSILRINRNRCQYCRLKKCIAVGMSRDAVR
H.virescens         GCK.....GFFRRSIQQKIQYRPCTKNQQCSILRINRNRCQYCRLKKCIAVGMSRDAVR

D.melanogaster(A)   FGRVPKREKARILAAMQQSTQNRGQQRALATELDDQPRLLAAVLRAHLETCEFTKEKVSA
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   FGRVPKREKARILAAMQQSTQNRGQQRALATELDDQPRLLAAVLRAHLETCEFTKEKVSA
M.sexta(B)          FGRVPKREKARILAAMQQSSTSRAHEQAAAAELDDAPRLLARVVRAHLDTCEFTRDRVAA
D.melanogaster(C)   FGRVPKREKARIWRPCNRAPRIAASSDPSPPSWMTSHASSPPCCCAHLETCEFTKEKVSA
C.fumiferana        FGRVPKREKARILAAMQQSSSSRAHEQAAAAELDDAPRLLARVVRAHLDTCEFTRDRVAA
G.mellonella        FGRVPKREKARILAAM.QSSTTRAHEQAAAAELDDGPRLLARVVRAHLDTCEFTRDRVAA
M.ensis             FGRVPKREKAKILAAM.QSVNAKSQERAVLAELEDDTRVTAAIIRAHMDTCDFTRDKVAP
H.virescens         FGRVPKREKARILAAMQQSSTSRANEQAAAAELDDAPRLLARVVRAHLDTCEFTRDRVAA

D.melanogaster(A)   MRQRARDCPSYSM.PTLLACPLNPAP.ELQSEQE....FSQRFAHVIRGVIDFAGMIPGF
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   MRQRARDCPSYSM.PTLLACPLNPAP.ELQSEQE....FSQRFAHVIRGVIDFAGMIPGF
M.sexta(B)          MRARARDCPTYSQ.PT.LACPLNPAP.ELQSEKE....FSQRFAHVIRGVIDFAGLIPGF
D.melanogaster(C)   MR.HGRGLPSTPC.HTS.GLSAEPAP.ELQSEQE....FSQRFAHVIRGVIDFAGMIPGF
C.fumiferana        MRARARDCPTYSQ.PT.LACPLNPAP.ELQSEKE....FSQRFAHVIRGVIDFAGLIPGF
G.mellonella        MRNGARDCPTYSQ.PT.LACPLNPAP.ELQSEKE....FSQRFAHVIRGVIDFAGLIPGF
M.ensis             MLQQARTHPSYTQCPPYLACPLNPRPVPLHGQQELVQDFSEALLPAIRGVVEFAKRLPGF
H.virescens         MRARARDCPIYSQ.PT.LACPLNPAP.ELQSEKE....FSQRFAHVIRGVIDFAGLIPGF

D.melanogaster(A)   QLLTQDDKFTLLKAGLFDALFVRLICMFDSSINSIICLNGQVMRRDAIQNGANARFLVDS
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   QLLTQDDKFTLLKAGLFDALFVRLICMFDSSINSIICLNGQVMRRDAIQNGANARFLVDS
M.sexta(B)          QLLTQDDKFTLLKSGLFDALFVRLICMFDAPLNSIICLNGQLMKRDSIQSGANARFLVDS
D.melanogaster(C)   QLLTQDDKFTLLKAGLFDALFVRLICMFDSSINSIICLNGQVMRRDAIQNGANARFLVDS
C.fumiferana        QLLTQDDKFTLLKSGLFDALFVRLICMFDAPLNSIICLNGQLMKRDSIQSGANARFLVDS
G.mellonella        QLLTQDDKFTLLKSGLFDALFVRLICMFDAPLNSIICLNGQLMKRDSIQSGANARFLVDS
M.ensis             QQLPQEDQVTLLKAGVFEVLLVRLAGMFDARTNAMLCLNGQLVRREALHTSVNARFLMDS
H.virescens         QLLTQDDKFTLLKSGLFDALFVRLICMFDAPLNSIICLNGQVMKRDSIQSGANARFLVDS
```

Figure 4B

```
D.melanogaster(A)   TFNFAERMNSMNLTDAEIGLFCAIVLITPDRPGLRNLELIEKMYSRLKGCLQYIVAQNRP
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   TFNFAERMNSMNLTDAEIGLFCAIVLITPDRPGLRNLELIEKMYSRLKGCLQYIVAQNRP
M.sexta(B)          TFKFAERMNSMNLTDAEIGLFCAIVLITPDRPGLRNVELVERMHTRLKACLQTVIAQNRP
D.melanogaster(C)   TFNFAERMNSMNLTDAEIGLFCAIVLITPDRPGLRNLELIEKMYSRLKGCLQYIVAQNRP
C.fumiferana        TFKFAERMNSMNLTDAEIGLFCAIVLITPDRPGLRNIELVERMHARLKSCLQTVIAQNRA
G.mellonella        TFKFAERMNSMNLTDAEIGLFCAIVLITPDRPGLRNVELVERMHSRLKSCLQTVIAQNRS
M.ensis             MFDFAERVNSLALNDAELALFCAVVVLAPDRPGLRNAELVERVHRRLVNCLQAVVSKHHP
H.virescens         TFKFAERMNSMNLTDAEIALFCAIVLITPDRPGLRNVELVERMHARLKACLQTVVAQNRP

D.melanogaster(A)   DQPEFLAKLLETMPDLRTLSTLHTEKLVVFR.TEHKELLRQQMWSMEDGNNSDGQQNKSP
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   DQPEFLAKLLETMPDLRTLSTLHTEKLVVFR.TEHKELLRQQMWSMEDGNNSDGQQNKSP
M.sexta(B)          DRPGFLRELMDTLPDLRTLSTLHTEKLVVFR.TEHKELLRQQMWSEEE............
D.melanogaster(C)   DQPEFLAKLLETMPDLRTLSTLHTEKLVVFR.TEHKELLRQQMWSMEDGNNSDGQQNKSP
C.fumiferana        DRPGFLRELMDTLPDLRTLSTLHTEKLVVFR.TEHKELLRQQMWGDEEV...........
G.mellonella        DGPGFLRELMDTLPDLRTLSTLHTEKLVVFR.TEHKELLRQQMWVEDEG...........
M.ensis             ENPNLQRDLLSKIPDLRTLNTLHSEKLLKYKMTEH..TAAGAPWDDSRSSWSMEQESSVG
H.virescens         DRPGFLRELMDTLPDLRTLSTLHTEKLVVFR.TEHKELLRQQMWTDEEG...........

D.melanogaster(A)   SGSWADAMDVEAAKSPLG...SVSSTESADLDYGSPSSSQPQGVSLPSPPQQQPSALASS
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   SGSWADAMDVEAAKSPLG...SVSSTESADLDYGSPSSSQPQGVSLPSPPQQQPSALASS
M.sexta(B)          AVSWVDSGADELARSPIG...SVSSSESGE.....................AVGDCG
D.melanogaster(C)   SGSWADAMDVEAAKSPLG...SVSSTESADLDYGSPSSSQPQGVSLPSPPQQQPSALASS
C.fumiferana        CP.WADSGVDDSARSPLG...SVSSSESGE.....................APSDCG
G.mellonella        AL.WADSGADDSARSPIG...SVSSSESSE.....................TTGDCG
M.ensis             SPS.SSYTTDEAMRSPVSCSESICSGESASSGESLCGSEVSGYTELRPPFPLARRRHDHS
H.virescens         VMSWGDSGADESARSPIG...SVSSSESGE.....................AVGDCG

D.melanogaster(A)   APLLAATLSGGCPLRNRANSGSSGDSGAAEMDIVGSHA.HLTQNGLTITPIVRHQQQQQQ
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   APLLAATLSGGCPLRNRANSGSSGDSGAAEMDIVGSHA.HLTQNGLTITPIVRHQQQQQQ
M.sexta(B)          TPLLAATLAG....RRRLDSRGSVDEEALGV......A.HLAHNGLTVTPV.........
D.melanogaster(C)   APLLAATLSGGCPLRNRANSGSSGDSGAAEMDIVGSHA.HLTQNGLTITPIVRHQQQQQQ
C.fumiferana        TPLLAATLAG....RRRLDSRGSVDEEALGV......A.HLAHNGLTVTPV.........
G.mellonella        TPLLAATLAG....RRRLDSRGSVDEEALGV......A.HLAHNGLTVTPV.........
M.ensis             EGASSGDEATESPLKCPFSKRKSDSPDDSGIESGTDRSDKLSSPSVCSSPRSSIDEKERG
H.virescens         TPLLAATLAG....RRRLDSRGSVDEEALGV......A.HLAHNGLTVTPV.........

D.melanogaster(A)   QQQIGILNNAHSRNLNGGHAMCQQQQQHPQLHHHLTAGAARYRKLDSPTDSGIESGNEK.
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   QQQIGILNNAHSRNLNGGHAMCQQQQQHPQLHHHLTAGAARYRKLDSPTDSGIESGNEK.
M.sexta(B)          ..............................RQPPRYRKLDSPTDSGIESGNEK.
D.melanogaster(C)   QQQIGILNNAHSRNLNGGHAMCQQQQQHPQLHHHLTAGAARYRKLDSPTDSGIESGNEK.
C.fumiferana        ..............................RPPPRYRKLDSPTDSGIESGNEK.
G.mellonella        ..............................RPPPRYRKLDSPTDSGIESGNEK.
M.ensis             GPARTICRCCARLQRRPSSTRICSWRKPTTSPIKSSVRNVGKRSLTPHSPPPPRSWSRRC
H.virescens         ..............................RPPPRYRKLDSPTDSGIESGNEK.

D.melanogaster(A)   ...NECKAVSSGGSSSCSSP.RSSVDDALDCSDAAANHNQVVQHPQLSVVSVSPVRSPQP
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ...NECKAVSSGGSSSCSSP.RSSVDDALDCSDAAANHNQVVQHPQLSVVSVSPVRSPQP
M.sexta(B)          ...HE.RIV..GTGSGCSSP.RSSLEEHNE.....................DRRPPV
D.melanogaster(C)   ...NECKAVSSGGSSSCSSP.RSSVDDALDCSDAAANHNQVVQHPQLSVVSVSPVRSPQP
C.fumiferana        ...HE.RIV..GPGSGCSSP.RSSLEEHME.....................DRRPLA
G.mellonella        ...HE.RIV..GPESGCSSP.RSSLEEHSD.....................DRRPIA
M.ensis             LSLHSTRALWLRHTPPWPPVWRRPLA~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ...HE.RIV..GPGSGCSSP.RSSLEEHTD.....................DRRPPP
```

Figure 4C

```
D.melanogaster(A)   STSSHLKRQIVEDMPVLKRVLQAPPLY.DTNSLMDEAYKPHKKFRALRHREFETAEADAS
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   STSSHLKRQIVEDMPVLKRVLQAPPLY.DTNSLMDEAYKPHKKFRALRHREFETAEADAS
M.sexta(B)          S.........ADDMPVLKRVLQAPPLYGGTPSLMDEAYRRHKKFRALRRDT.GEAEA...
D.melanogaster(C)   STSSHLKRQIVEDMPVLKRVLQAPPLY.DTNSLMDEAYKPHKKFRALRHREFETAEADAS
C.fumiferana        ..........ADDMPVLKRVLQAPPLY.DASSLMDEAYKPHKKFRAMRRDT.GEAEA...
G.mellonella        P.........ADDMPVLKRVLQAPPLY.DASSLMDEAYKPHKKFRAMRRDTWSEAEA...
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         S.........ADDMPVLKRVLEAPPLY.HTTSLMDEAYKPHKKFRAMRRDT.GEAEA...

D.melanogaster(A)   SSTSGSNSLSAGSPRQSPVPNSVATPPPSAASAAAGNPAQSQLHMHLTRSSPKASMASSH
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   SSTSGSNSLSAGSPRQSPVPNSVATPPPSAASAAAGNPAQSQLHMHLTRSSPKASMASSH
M.sexta(B)          .........RTVRPTPSPQP...QHPHP....ANPAHPAHSPRP.......QRASLSSTH
D.melanogaster(C)   SSTSGSNSLSAGSPRQSPVPNSVATPPPVAASAAAGNPAQSQLHMHLTRSSPKASMASSH
C.fumiferana        .........RPMRPTPSPQPMHPHPGSP....AHPAHPAHSPRPL.......RAPLSSTH
G.mellonella        .........RPGRPTPSPQP..PHHPHP....ASPAHPAHSPRPI.......RAPLSSTH
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         .........RVVRPAPSTQP..PQHPHP....ASPAHPAHSPRPL.......RASLSSTH

D.melanogaster(A)   SVLAKSLMAEPRMTPEQMKRSDIIQNYLKRENSTAASS..TTNGVGNRSPSSSSTPPPSA
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   SVLAKSLMAEPRMTPEQMKRSDIIQNYLKRENSTAASS..TTNGVGNRSPSSSSTPPPSA
M.sexta(B)          SVLAKSLMEGPRMTPEQLKRTDIIQQYMRRGESSAPAEGCPLRAGGLLTCYRGASPAPQP
D.melanogaster(C)   SVLAKSLMAEPRMTPEQMKRSDIIQNYLKRENSTAASS..TTNGLGNRSPSSSSTPPPS.
C.fumiferana        SVLAKSLMEGPRMTPEQLKRTDIIQQYMRRGEAGE.....ECRAGLLL..YRGASP....
G.mellonella        SVLAKSLMEGPRMTPEQLKRTDIIQQYMRRGETGAPTEGCPLRAGGLLTCFRGASPAPQP
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         SVLAKSLMEGPRMTPEQLKRTDIIQQYMRRGEAGAP.DGCPMRTGGLLTCYRGASPAPQP

D.melanogaster(A)   VQNQQRWGSSSVITTTCQQRQQSVSPHSNGSSSSSSSSSSSSSSSSTSSNCSSSSASSC
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   VQNQQRWGSSSVITTTCQQRQQSVSPHSNGSSSSSSSSSSSSSSSSTSSNCSSSSASSC
M.sexta(B)          VLALQVDVTDA....PLNLSKKSPSP......PRTYMPQMLEA~~~~~~~~~~~~~~~~~
D.melanogaster(C)   VQNQQRWGSSSVITTTCQQRQQSVSPHSNGSSSSSSSSSSSSSSSSTSSNCSSSSASSC
C.fumiferana        ...LQVDVA..DAPQPLNLSKKSPS......PPRSFMPPMLEA~~~~~~~~~~~~~~~~~
G.mellonella        VIALQVDVAETDAPQPLNLSKKSPSPSPPPPPPRSYMPPMLPA~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         VMALQVDVSDADA..PLNLSKKSPS......PPRSFMPQMLEA~~~~~~~~~~~~~~~~~

D.melanogaster(A)   QYFQSPHSTSNGTSAPASSSSGSNSATPLLELQVDIADSAQPLNLSKKSPTPPPSKLHAL
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   QYFQSPHSTSNGTSAPASSSSGSNSATPLLELQVDIADSAQPLNLSKKSPTPPPSKLHAL
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   QYFQSPHSTSIGTGEPDGAPVRDRTAPRPCWNCRWTLLTRRTSQFVQEIAHAAAQQAARS
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   VAAANAVQRYPTLSADVTVTASNGGSSVGGGESGRQQQSAGECGLPQSGPERRRAQGNAG
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   VAAANAVQRYPTLSADVTVTASNGGSSVGGGESGRQQQSAGECGLPQSGPERRRAQGNAG
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   GGRRQCRSKVSHIVRRRHSDSLQWRSSVGGGESGAQQQSAGECGLPQSGPERRRAQGNAG
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Figure 4D

```
D.melanogaster(A)   GVRAGGGRWFYAEKWERQRLGVAVQRSRKQDHLERRELN*IILPFN*DVYKV*KQNQHAC
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   GVRAGGGRWFYAEKWERQRLGVAVQRSRKQDHLERRELN~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   GVRAGGGRWFYAEKWERQRLGVAVQRSRKQDHLERRELN~~~~~~~~~~~~~~~~~~~~~
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   NLKLIFKATTNKTTTSY*FKKQTNKQTTKNPSLNGITKEKEKQKKYKYILAVKL*RSKKP
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   TNPRQRSDFALTFLQLLPKTPLTSPPPNPSSTHQPSFDP*LFYKF*ALVVHINYVYW*LC
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   LAL*L*LEQNYFAFLDVF*KNCKLLLLNF*IPKNKTMCVKFFIVRSPSRMKCSLQQILTT
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   IKLITIHFL*I*LIL*ICYSFPPFYRSFYLII*LPVFLISPLAQSSSLC*RIKWNKYCFL
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   ILKLPQKYD*NIHEVIENQTKCLKF*QQAVKRR*RRETQR*IYLLCT*LNVKLNTKTYLK
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Figure 4E

```
D.melanogaster(A)   YI*IHIIINEETYA*KIQCLIGILENQAKNTKKNQQTKIMIYYLKVKYTFTLQKNKRENL
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   R*QQNYYINYILIMLYYYDY*LL*LINYDFYA*TNQQKTNMQKPLKKKTKNKQKITLAQK
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   FVLSNQEKICLKISKESLLSFFISISFSVEHFFLKCSVLPLLCFGVVRFWVSLDFIFVLT
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

D.melanogaster(A)   IVL*KYLLNYIESIYIIIK*TR~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(A)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(B)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.sexta(B)          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
D.melanogaster(C)   ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
C.fumiferana        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
G.mellonella        ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
M.ensis             ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
H.virescens         ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

Figure 4F

FIG. 5

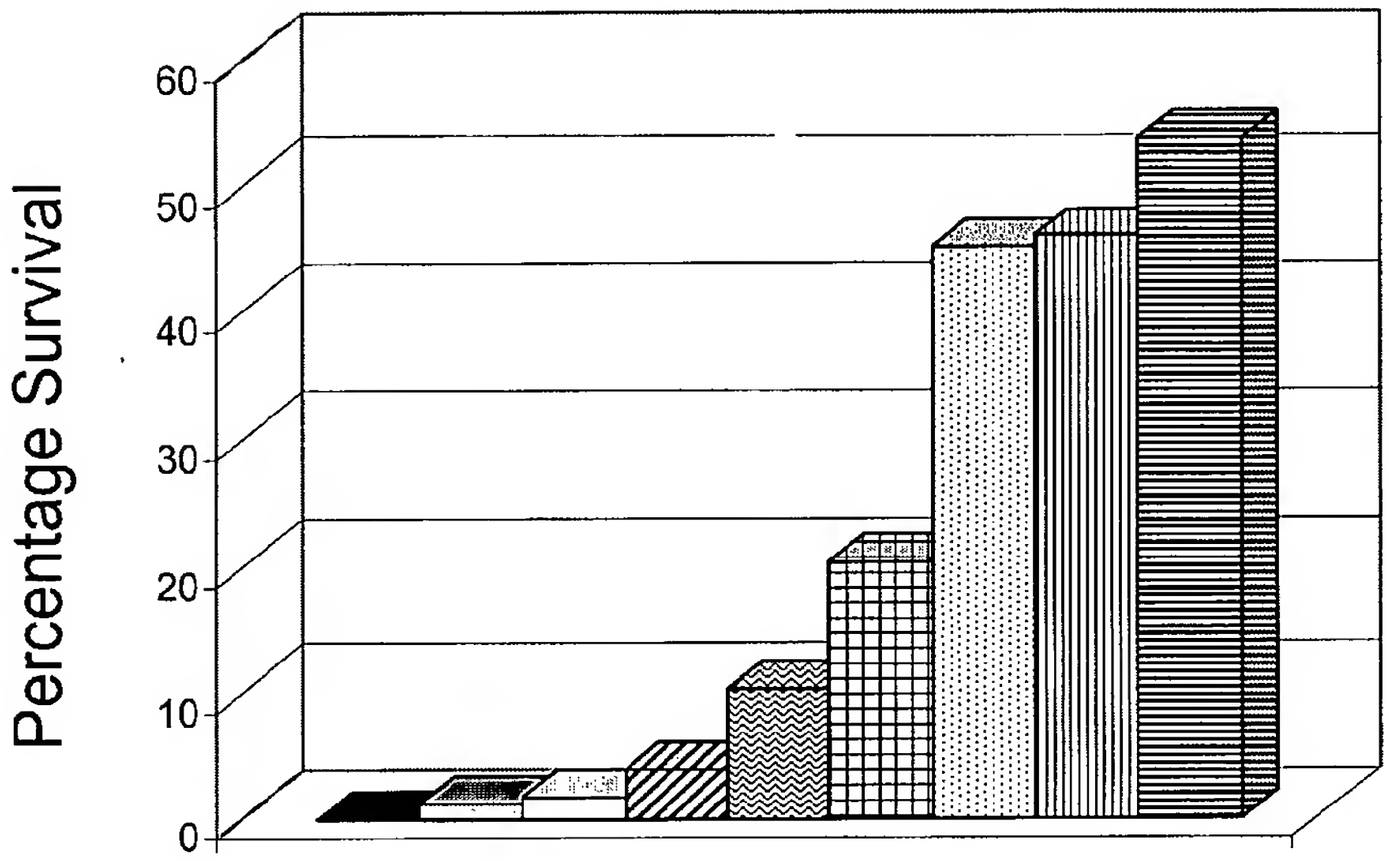
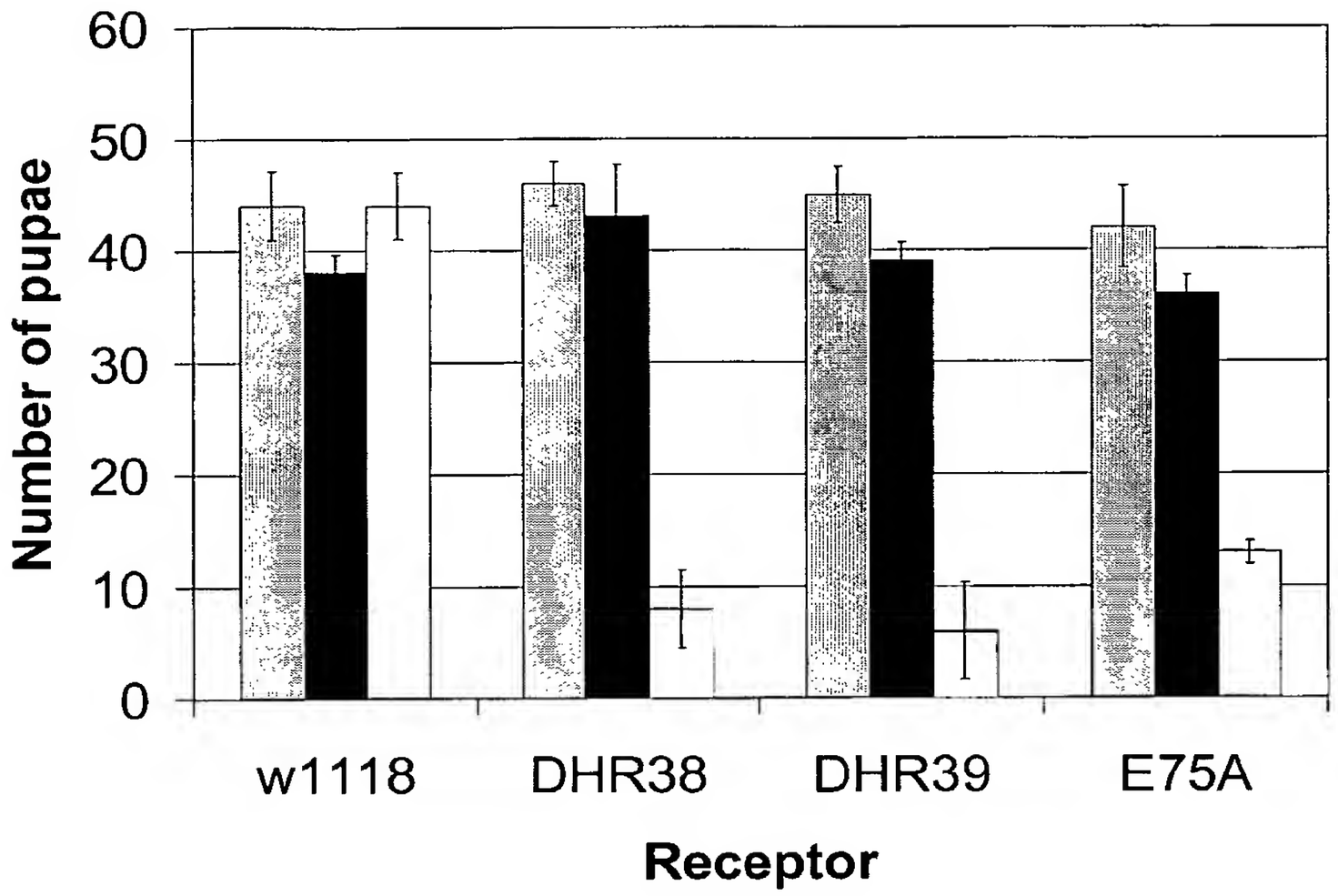


FIG. 6


60
50
40
30
20
10
0
Number of pupae
w1118
DHR38
DHR39
E75A
Receptor

Sequence Listing

<110> Syngenta
Broadus, Julie
Brown, Blanche
Stam, Lynn
Kamdar, Kim

<120> INSECT NUCLEAR RECEPTOR GENES AND USES THEREOF

<130> 1392/3

<150> US 60/278,336

<151> 2001-03-23

<160> 56

<170> PatentIn version 3.0

<210> 1
<211> 1560
<212> DNA
<213> Drosophila melanogaster
<400> 1
atgctgagga acgaatacat aagcaataac aagaaggtgc tcaactcgga ccaaaacaag 60

tactacatgc taacggtcga ggaggccgat atccaaggca cgaacatgtc cgacggcgtc 120

agcatcttgc acatcaaaca ggaggtggac actccatcgg cgtcctgctt tagtcccagc 180

tccaagtcaa cggccacgca gagtggcaca aacggcctga aatcctcgcc ctcggtttcg 240

ccggaaaggc agctctgcag ctcgacgacc tctctatcct gcgatttgca caatgtatcc 300

ttaagcaatg atggcgatag tctgaaagga agtggtacaa gtggcggcaa tggcggagga 360

ggaggtggtg gtacgagtgg tggaaatgcg accaatgcga gtgccggagc tggatcggga 420

tccgtcaggg acgagctccg ccgattgtgt ttggtttgtg gcgatgtggc cagtggattc 480

cactatggtg tggcgagttg tgaggcttgc aaagcgttct ttaaacgcac catccaaggc 540

aacatcgagt acacgtgtcc ggcgaacaac gagtgtgaga ttaacaagcg gagacgcaag 600

gcctgccaag cgtgtcgctt ccagaaatgt ctactaatgg gcatgctcaa ggagggtgtg 660

cgcttggatc gagttcgtgg aggacggcag aagtaccgaa ggaatcctgt atcaaactct 720

taccagacta tgcagctgct ataccaatcc aacaccacct cgctgtgcga tgtcaagata 780

ctggaggtgc tcaattcata tgagccggat gccttgagcg tccaaacgcc gccgccgcaa 840

gtccacacga ctagcataac taatgatgag gcctcatcct cctcgggcag cataaaactg 900

gagtccagcg ttgttacgcc caatgggact tgcattttcc aaaacaacaa caacaatgat 960

cccaatgaga tactaagcgt ccttagtgat atttacgaca aggaattggt cagcgtcatt 1020

ggctgggcca agcagatacc tggctttata gatctgccac ttaacgacca gatgaagctt 1080

ctccaggtgt cgtgggcaga gatcctgacg ctccagctga ccttccggtc cctaccgttc 1140

aatggcaagt tatgcttcgc cacggatgtc tggatggatg aacatttggc caaggagtgc 1200

ggttacacgg agttctacta ccactgcgtc cagatcgcac agcgcatgga aagaatatcg 1260

ccacgaaggg aggagtacta cttgctaaag gcgctcctgc tggccaactg cgacattctg 1320

ctggatgatc agagttccct gcgcgcattt cgtgatacga ttcttaattc tctaaacgat 1380

gtggtctact tgctgcgtca ttcgtcggcc gtgtcgcatc agcaacaatt gctgcttttg 1440

ctgccttcgc tgcggcaggc ggatgatatc ctgcgaagat tttggcgtgg aattgcacgc 1500

gatgaagtca ttaccatgaa gaaactgttc ctcgagatgc tcgagccgct ggccaggtga 1560

<210> 2
<211> 519
<212> PRT
<213> Drosophila melanogaster
<400> 2

Met Leu Arg Asn Glu Tyr Ile Ser Asn Asn Lys Lys Val Leu Asn Ser
1 5 10 15

Asp Gln Asn Lys Tyr Tyr Met Leu Thr Val Glu Glu Ala Asp Ile Gln
20 25 30

Gly Thr Asn Met Ser Asp Gly Val Ser Ile Leu His Ile Lys Gln Glu
35 40 45

Val Asp Thr Pro Ser Ala Ser Cys Phe Ser Pro Ser Ser Lys Ser Thr
50 55 60

Ala Thr Gln Ser Gly Thr Asn Gly Leu Lys Ser Ser Pro Ser Val Ser
65 70 75 80

Pro Glu Arg Gln Leu Cys Ser Ser Thr Thr Ser Leu Ser Cys Asp Leu
85 90 95

His Asn Val Ser Leu Ser Asn Asp Gly Asp Ser Leu Lys Gly Ser Gly
100 105 110

Thr Ser Gly Gly Asn Gly Gly Gly Gly Gly Gly Gly Thr Ser Gly Gly
115 120 125

Asn Ala Thr Asn Ala Ser Ala Gly Ala Gly Ser Gly Ser Val Arg Asp
130 135 140

Glu Leu Arg Arg Leu Cys Leu Val Cys Gly Asp Val Ala Ser Gly Phe
145 150 155 160

His Tyr Gly Val Ala Ser Cys Glu Ala Cys Lys Ala Phe Phe Lys Arg
165 170 175

Thr Ile Gln Gly Asn Ile Glu Tyr Thr Cys Pro Ala Asn Asn Glu Cys
180 185 190

```
Glu Ile Asn Lys Arg Arg Arg Lys Ala Cys Gln Ala Cys Arg Phe Gln
        195                 200                 205

Lys Cys Leu Leu Met Gly Met Leu Lys Glu Gly Val Arg Leu Asp Arg
    210                 215                 220

Val Arg Gly Gly Arg Gln Lys Tyr Arg Arg Asn Pro Val Ser Asn Ser
225                 230                 235                 240

Tyr Gln Thr Met Gln Leu Leu Tyr Gln Ser Asn Thr Thr Ser Leu Cys
                245                 250                 255

Asp Val Lys Ile Leu Glu Val Leu Asn Ser Tyr Glu Pro Asp Ala Leu
            260                 265                 270

Ser Val Gln Thr Pro Pro Pro Gln Val His Thr Thr Ser Ile Thr Asn
        275                 280                 285

Asp Glu Ala Ser Ser Ser Ser Gly Ser Ile Lys Leu Glu Ser Ser Val
    290                 295                 300

Val Thr Pro Asn Gly Thr Cys Ile Phe Gln Asn Asn Asn Asn Asp
305                 310                 315                 320

Pro Asn Glu Ile Leu Ser Val Leu Ser Asp Ile Tyr Asp Lys Glu Leu
                325                 330                 335

Val Ser Val Ile Gly Trp Ala Lys Gln Ile Pro Gly Phe Ile Asp Leu
            340                 345                 350

Pro Leu Asn Asp Gln Met Lys Leu Leu Gln Val Ser Trp Ala Glu Ile
        355                 360                 365

Leu Thr Leu Gln Leu Thr Phe Arg Ser Leu Pro Phe Asn Gly Lys Leu
    370                 375                 380

Cys Phe Ala Thr Asp Val Trp Met Asp Glu His Leu Ala Lys Glu Cys
385                 390                 395                 400

Gly Tyr Thr Glu Phe Tyr Tyr His Cys Val Gln Ile Ala Gln Arg Met
                405                 410                 415

Glu Arg Ile Ser Pro Arg Arg Glu Glu Tyr Tyr Leu Leu Lys Ala Leu
            420                 425                 430

Leu Leu Ala Asn Cys Asp Ile Leu Leu Asp Asp Gln Ser Ser Leu Arg
        435                 440                 445

Ala Phe Arg Asp Thr Ile Leu Asn Ser Leu Asn Asp Val Val Tyr Leu
    450                 455                 460

Leu Arg His Ser Ser Ala Val Ser His Gln Gln Gln Leu Leu Leu Leu
465                 470                 475                 480

Leu Pro Ser Leu Arg Gln Ala Asp Asp Ile Leu Arg Arg Phe Trp Arg
                485                 490                 495

Gly Ile Ala Arg Asp Glu Val Ile Thr Met Lys Lys Leu Phe Leu Glu
            500                 505                 510
```

Met Leu Glu Pro Leu Ala Arg
515

<210> 3
<211> 1455
<212> DNA
<213> Drosophila melanogaster
<400> 3

atgtccgacg gcgtcagcat cttgcacatc aaacaggagg tggacactcc atcggcgtcc 60

tgctttagtc ccagctccaa gtcaacggcc acgcagagtg gcacaaacgg cctgaaatcc 120

tcgccctcgg tttcgccgga aaggcagctc tgcagctcga cgacctctct atcctgcgat 180

ttgcacaatg tatccttaag caatgatggc gatagtctga aaggaagtgg tacaagtggc 240

ggcaatggcg gaggaggagg tggtggtacg agtggtggaa atgcgaccaa tgcgagtgcc 300

ggagctggat cgggatccgt cagggacgag ctccgccgat tgtgtttggt ttgtggcgat 360

gtggccagtg gattccacta tggtgtggcg agttgtgagg cttgcaaagc gttctttaaa 420

cgcaccatcc aaggcaacat cgagtacacg tgtccggcga acaacgagtg tgagattaac 480

aagcggagac gcaaggcctg ccaagcgtgt cgcttccaga aatgtctact aatgggcatg 540

ctcaaggagg gtgtgcgctt ggatcgagtt cgtggaggac ggcagaagta ccgaaggaat 600

cctgtatcaa actcttacca gactatgcag ctgctatacc aatccaacac cacctcgctg 660

tgcgatgtca agatactgga ggtgctcaat tcatatgagc cggatgcctt gagcgtccaa 720

acgccgccgc cgcaagtcca cacgactagc ataactaatg atgaggcctc atcctcctcg 780

ggcagcataa aactggagtc cagcgttgtt acgcccaatg ggacttgcat tttccaaaac 840

aacaacaaca atgatcccaa tgagatacta agcgtcctta gtgatattta cgacaaggaa 900

ttggtcagcg tcattggctg ggccaagcag atacctggct ttatagatct gccacttaac 960

gaccagatga agcttctcca ggtgtcgtgg gcagagatcc tgacgctcca gctgaccttc 1020

cggtccctac cgttcaatgg caagttatgc ttcgccacgg atgtctggat ggatgaacat 1080

ttggccaagg agtgcggtta cacggagttc tactaccact gcgtccagat cgcacagcgc 1140

atggaaagaa tatcgccacg aagggaggag tactacttgc taaaggcgct cctgctggcc 1200

aactgcgaca ttctgctgga tgatcagagt tccctgcgcg catttcgtga tacgattctt 1260

aattctctaa acgatgtggt ctacttgctg cgtcattcgt cggccgtgtc gcatcagcaa 1320

caattgctgc ttttgctgcc ttcgctgcgg caggcggatg atatcctgcg aagattttgg 1380

cgtggaattg cacgcgatga agtcattacc atgaagaaac tgttcctcga gatgctcgag 1440

ccgctggcca ggtga 1455

<210> 4
<211> 484
<212> PRT
<213> Drosophila melanogaster
<400> 4

Met Ser Asp Gly Val Ser Ile Leu His Ile Lys Gln Glu Val Asp Thr
1 5 10 15

Pro Ser Ala Ser Cys Phe Ser Pro Ser Ser Lys Ser Thr Ala Thr Gln
20 25 30

Ser Gly Thr Asn Gly Leu Lys Ser Ser Pro Ser Val Ser Pro Glu Arg
35 40 45

Gln Leu Cys Ser Ser Thr Thr Ser Leu Ser Cys Asp Leu His Asn Val
50 55 60

Ser Leu Ser Asn Asp Gly Asp Ser Leu Lys Gly Ser Gly Thr Ser Gly
65 70 75 80

Gly Asn Gly Gly Gly Gly Gly Gly Gly Gly Thr Ser Gly Gly Asn Ala Thr
85 90 95

Asn Ala Ser Ala Gly Ala Gly Ser Gly Ser Val Arg Asp Glu Leu Arg
100 105 110

Arg Leu Cys Leu Val Cys Gly Asp Val Ala Ser Gly Phe His Tyr Gly
115 120 125

Val Ala Ser Cys Glu Ala Cys Lys Ala Phe Phe Lys Arg Thr Ile Gln
130 135 140

Gly Asn Ile Glu Tyr Thr Cys Pro Ala Asn Asn Glu Cys Glu Ile Asn
145 150 155 160

Lys Arg Arg Arg Lys Ala Cys Gln Ala Cys Arg Phe Gln Lys Cys Leu
165 170 175

Leu Met Gly Met Leu Lys Glu Gly Val Arg Leu Asp Arg Val Arg Gly
180 185 190

Gly Arg Gln Lys Tyr Arg Arg Asn Pro Val Ser Asn Ser Tyr Gln Thr
195 200 205

Met Gln Leu Leu Tyr Gln Ser Asn Thr Thr Ser Leu Cys Asp Val Lys
210 215 220

Ile Leu Glu Val Leu Asn Ser Tyr Glu Pro Asp Ala Leu Ser Val Gln
225 230 235 240

Thr Pro Pro Pro Gln Val His Thr Thr Ser Ile Thr Asn Asp Glu Ala
245 250 255

Ser Ser Ser Ser Gly Ser Ile Lys Leu Glu Ser Ser Val Val Thr Pro
260 265 270

Asn Gly Thr Cys Ile Phe Gln Asn Asn Asn Asn Asn Asp Pro Asn Glu
275 280 285

Ile Leu Ser Val Leu Ser Asp Ile Tyr Asp Lys Glu Leu Val Ser Val
290 295 300

Ile Gly Trp Ala Lys Gln Ile Pro Gly Phe Ile Asp Leu Pro Leu Asn
305 310 315 320

Asp Gln Met Lys Leu Leu Gln Val Ser Trp Ala Glu Ile Leu Thr Leu
325 330 335

Gln Leu Thr Phe Arg Ser Leu Pro Phe Asn Gly Lys Leu Cys Phe Ala
340 345 350

Thr Asp Val Trp Met Asp Glu His Leu Ala Lys Glu Cys Gly Tyr Thr
355 360 365

Glu Phe Tyr Tyr His Cys Val Gln Ile Ala Gln Arg Met Glu Arg Ile
370 375 380

Ser Pro Arg Arg Glu Glu Tyr Tyr Leu Leu Lys Ala Leu Leu Leu Ala
385 390 395 400

Asn Cys Asp Ile Leu Leu Asp Asp Gln Ser Ser Leu Arg Ala Phe Arg
405 410 415

Asp Thr Ile Leu Asn Ser Leu Asn Asp Val Val Tyr Leu Leu Arg His
420 425 430

Ser Ser Ala Val Ser His Gln Gln Gln Leu Leu Leu Leu Leu Pro Ser
435 440 445

Leu Arg Gln Ala Asp Asp Ile Leu Arg Arg Phe Trp Arg Gly Ile Ala
450 455 460

Arg Asp Glu Val Ile Thr Met Lys Lys Leu Phe Leu Glu Met Leu Glu
465 470 475 480

Pro Leu Ala Arg

<210> 5
<211> 2106
<212> DNA
<213> Drosophila melanogaster
<400> 5
atgaccactg tcaaggcgga gaagccggta accaccactt gctcctactg ggtgttacgc 60

aaaaagcgcc gctgcaagat gactgcaaac aagggcagtg agttttgtgg agcccatgcc 120

tcatcggcag ccaccactgc aacttgcgat gaaaaagcta gagaagattc cttccaggaa 180

cgtattccct gcccgcttga tcacaaacat accgtgttta aaagaaaact agccaaacac 240

ttgaccattt gcaatgccag ggatcaggaa agttcactgc cctacattgt aaaaggagtg 300

aattctggag acaatcttaa ggaaacagac gaagatttga acaaattcaa tcaaataaaa 360

ctgcacgaac tggcggatga agagttctac agtctgatcg acaaagttaa gaatctgtat 420

gacaagcaca tcaattccag catacaggaa ctgcagctgg agcacgaatc cctaaaggag 480

gaacttagtc gcaaggacta cggccaggaa acgctacgcc agctcaccca gaccactagc 540

ttgctgggca tcctggaaca cgatcaccag ttgatggatc acacgagcta tatagaattc 600

ggagccggaa agggacaatt ggcctacttt ttggccaccg tattgcagga acagaagttg 660

agtcactcac aggtggtcct catcgatcga atgtccctgc ggcacaagaa ggacaacaag 720

ttggccaaca gggaggtggt gcaacgtatc cgtgctgata tcgctgatct taagctttct 780

gctctgccgg agctaaagaa aacccaacga acggtggctt tttcgaagca cctctgtgga 840

gcagctacag acttgaccct gagatgtata ctcggtgatg gaaatgccag ctcggactac 900

gttctcatcg ccctgtgttg tcatcatcgc tgctcctggc gctcctacgt gggacgcaag 960

tttcttcagg aagctggaat tggaccacgg gagttcgtta tcctaaccaa aatggtcagt 1020

tgggcagttt gtggcacggg cttaagccga gagcggcgta aggccatgga atcggctgac 1080

tttcagctca ccgagacgaa cacgcagcgt ctaactcgcc aggagcgcga acaaattggc 1140

caacagtgca agcgggttct ggattatgga cgactggagc atctacgatc ccacggatat 1200

caagcagaac ttaagttcta tgttcccagg gatgtgactt tggaaaacca catccaggtg 1260

gcatataaaa ccgccaatat ccgggatgag ctgattcagt ccgcggcgac tctttggctc 1320

attggcaagc agtgccaacc ggtgaccatg tcgaacttca gtgcctgcgc agtgtgcggc 1380

gatcagagct ccgggaagca ctacggcgtg tcctgctgcg atgggtgctc ctgcttttc 1440

aagcggagcg tgcggcgcgg gagcagctac gcctgcatcg ctctggtcgg gaactgtgtg 1500

gtggacaagg cgcggcggaa ctggtgtccc tcctgccgct tccagcgatg cctggccgtg 1560

ggaatgaacg ctgctgcggt tcaggaggag cgcggtccgc gcaaccagca ggtggctctc 1620

taccgcactg gccggagaca agctccgcca tctcaggcgg cgccatcccc gacgccccac 1680

tcccaggcgc tgcacttcca gatcctcgcc cagatccttg tcacgtgcct gcgccaggcg 1740

aaggccaacg agcagttcgc tctgttggat cgctgccaac aagacgccat ctttcaggtg 1800

gtgtggagcg agatcttcgt cctgcgagcg tcccactggt ctctggacat cagcgccatg 1860

atcgacggct gcggcgatga gcagctcaaa cggctcattt gcgaggccca ccagctaagg 1920

gccgacgtcc tggaactcaa ctttatggag tccctaatcc tgtgcagaaa agaattggcc 1980

atcaatgcgg agtatgccgt tatcctggga agccactcta aagccgccct gatctcctta 2040

gcccgctaca ccctgcagca atccaactac ctgcgggtgg tcagggacat cttaaaaaca 2100

ctttag 2106

<210> 6
<211> 701
<212> PRT
<213> Drosophila melanogaster
<400> 6

```
Met Thr Thr Val Lys Ala Glu Lys Pro Val Thr Thr Thr Cys Ser Tyr
1               5                  10                  15

Trp Val Leu Arg Lys Lys Arg Arg Cys Lys Met Thr Ala Asn Lys Gly
            20                  25                  30

Ser Glu Phe Cys Gly Ala His Ala Ser Ser Ala Ala Thr Thr Ala Thr
        35                  40                  45

Cys Asp Glu Lys Ala Arg Glu Asp Ser Phe Gln Glu Arg Ile Pro Cys
    50                  55                  60

Pro Leu Asp His Lys His Thr Val Phe Lys Arg Lys Leu Ala Lys His
65                  70                  75                  80

Leu Thr Ile Cys Asn Ala Arg Asp Gln Glu Ser Ser Leu Pro Tyr Ile
                85                  90                  95

Val Lys Gly Val Asn Ser Gly Asp Asn Leu Lys Glu Thr Asp Glu Asp
            100                 105                 110

Leu Asn Lys Phe Asn Gln Ile Lys Leu His Glu Leu Ala Asp Glu Glu
        115                 120                 125

Phe Tyr Ser Leu Ile Asp Lys Val Lys Asn Leu Tyr Asp Lys His Ile
    130                 135                 140

Asn Ser Ser Ile Gln Glu Leu Gln Leu Glu His Glu Ser Leu Lys Glu
145                 150                 155                 160

Glu Leu Ser Arg Lys Asp Tyr Gly Gln Glu Thr Leu Arg Gln Leu Thr
                165                 170                 175

Gln Thr Thr Ser Leu Leu Gly Ile Leu Glu His Asp His Gln Leu Met
            180                 185                 190

Asp His Thr Ser Tyr Ile Glu Phe Gly Ala Gly Lys Gly Gln Leu Ala
        195                 200                 205

Tyr Phe Leu Ala Thr Val Leu Gln Glu Gln Lys Leu Ser His Ser Gln
    210                 215                 220

Val Val Leu Ile Asp Arg Met Ser Leu Arg His Lys Lys Asp Asn Lys
225                 230                 235                 240

Leu Ala Asn Arg Glu Val Val Gln Arg Ile Arg Ala Asp Ile Ala Asp
                245                 250                 255

Leu Lys Leu Ser Ala Leu Pro Glu Leu Lys Lys Thr Gln Arg Thr Val
            260                 265                 270

Ala Phe Ser Lys His Leu Cys Gly Ala Ala Thr Asp Leu Thr Leu Arg
        275                 280                 285

Cys Ile Leu Gly Asp Gly Asn Ala Ser Ser Asp Tyr Val Leu Ile Ala
    290                 295                 300

Leu Cys Cys His His Arg Cys Ser Trp Arg Ser Tyr Val Gly Arg Lys
305                 310                 315                 320
```

```
Phe Leu Gln Glu Ala Gly Ile Gly Pro Arg Glu Phe Val Ile Leu Thr
                325                 330                 335

Lys Met Val Ser Trp Ala Val Cys Gly Thr Gly Leu Ser Arg Glu Arg
            340                 345                 350

Arg Lys Ala Met Glu Ser Ala Asp Phe Gln Leu Thr Glu Thr Asn Thr
        355                 360                 365

Gln Arg Leu Thr Arg Gln Glu Arg Glu Gln Ile Gly Gln Gln Cys Lys
    370                 375                 380

Arg Val Leu Asp Tyr Gly Arg Leu Glu His Leu Arg Ser His Gly Tyr
385                 390                 395                 400

Gln Ala Glu Leu Lys Phe Tyr Val Pro Arg Asp Val Thr Leu Glu Asn
                405                 410                 415

His Ile Gln Val Ala Tyr Lys Thr Ala Asn Ile Arg Asp Glu Leu Ile
            420                 425                 430

Gln Ser Ala Ala Thr Leu Trp Leu Ile Gly Lys Gln Cys Gln Pro Val
        435                 440                 445

Thr Met Ser Asn Phe Ser Ala Cys Ala Val Cys Gly Asp Gln Ser Ser
    450                 455                 460

Gly Lys His Tyr Gly Val Ser Cys Cys Asp Gly Cys Ser Cys Phe Phe
465                 470                 475                 480

Lys Arg Ser Val Arg Arg Gly Ser Ser Tyr Ala Cys Ile Ala Leu Val
                485                 490                 495

Gly Asn Cys Val Val Asp Lys Ala Arg Arg Asn Trp Cys Pro Ser Cys
            500                 505                 510

Arg Phe Gln Arg Cys Leu Ala Val Gly Met Asn Ala Ala Ala Val Gln
        515                 520                 525

Glu Glu Arg Gly Pro Arg Asn Gln Gln Val Ala Leu Tyr Arg Thr Gly
    530                 535                 540

Arg Arg Gln Ala Pro Pro Ser Gln Ala Ala Pro Ser Pro Thr Pro His
545                 550                 555                 560

Ser Gln Ala Leu His Phe Gln Ile Leu Ala Gln Ile Leu Val Thr Cys
                565                 570                 575

Leu Arg Gln Ala Lys Ala Asn Glu Gln Phe Ala Leu Leu Asp Arg Cys
            580                 585                 590

Gln Gln Asp Ala Ile Phe Gln Val Val Trp Ser Glu Ile Phe Val Leu
        595                 600                 605

Arg Ala Ser His Trp Ser Leu Asp Ile Ser Ala Met Ile Asp Gly Cys
    610                 615                 620

Gly Asp Glu Gln Leu Lys Arg Leu Ile Cys Glu Ala His Gln Leu Arg
625                 630                 635                 640
```

Ala Asp Val Leu Glu Leu Asn Phe Met Glu Ser Leu Ile Leu Cys Arg
645 650 655

Lys Glu Leu Ala Ile Asn Ala Glu Tyr Ala Val Ile Leu Gly Ser His
660 665 670

Ser Lys Ala Ala Leu Ile Ser Leu Ala Arg Tyr Thr Leu Gln Gln Ser
675 680 685

Asn Tyr Leu Arg Val Val Arg Asp Ile Leu Lys Thr Leu
690 695 700

<210> 7
<211> 837
<212> DNA
<213> Drosophila melanogaster
<400> 7
atgtcgaact tcagtgcctg cgcagtgtgc ggcgatcaga gctccgggaa gcactacggc 60

gtgtcctgct gcgatgggtg ctcctgcttt ttcaagcgga gcgtgcggcg cgggagcagc 120

tacgcctgca tcgctctggt cgggaactgt gtggtggaca aggcgcggcg gaactggtgt 180

ccctcctgcc gcttccagcg atgcctggcc gtgggaatga acgctgctgc ggttcaggag 240

gagcgcggtc cgcgcaacca gcaggtggct ctctaccgca ctggccggag acaagctccg 300

ccatctcagg cggcgccatc cccgacgccc cactcccagg cgctgcactt ccagatcctc 360

gcccagatcc ttgtcacgtg cctgcgccag gcgaaggcca acgagcagtt cgctctgttg 420

gatcgctgcc aacaagacgc catctttcag gtggtgtgga gcgagatctt cgtcctgcga 480

gcgtcccact ggtctctgga catcagcgcc atgatcgacg gctgcggcga tgagcagctc 540

aaacggctca tttgcgaggc ccaccagcta agggccgacg tcctggaact caactttatg 600

gagtccctaa tcctgtgcag aaaagaattg gccatcaatg cggagtatgc cgttatcctg 660

ggaagccact ctaaagccgc cctgatctcc ttagcccgct acaccctgca gcaatccaac 720

tacctgcggt tcggacaact gctccttggt ctgaggcagc tgtgcctgag gcgcttcgac 780

tgcgcgcttt cttgtatgtt tcgcagcgtg gtcagggaca tcttaaaaac actttaa 837

<210> 8
<211> 278
<212> PRT
<213> Drosophila melanogaster
<400> 8

Met Ser Asn Phe Ser Ala Cys Ala Val Cys Gly Asp Gln Ser Ser Gly
1 5 10 15

Lys His Tyr Gly Val Ser Cys Cys Asp Gly Cys Ser Cys Phe Phe Lys
20 25 30

Arg Ser Val Arg Arg Gly Ser Ser Tyr Ala Cys Ile Ala Leu Val Gly
35 40 45

Asn Cys Val Val Asp Lys Ala Arg Arg Asn Trp Cys Pro Ser Cys Arg
50 55 60

Phe Gln Arg Cys Leu Ala Val Gly Met Asn Ala Ala Ala Val Gln Glu
65 70 75 80

Glu Arg Gly Pro Arg Asn Gln Gln Val Ala Leu Tyr Arg Thr Gly Arg
85 90 95

Arg Gln Ala Pro Pro Ser Gln Ala Ala Pro Ser Pro Thr Pro His Ser
100 105 110

Gln Ala Leu His Phe Gln Ile Leu Ala Gln Ile Leu Val Thr Cys Leu
115 120 125

Arg Gln Ala Lys Ala Asn Glu Gln Phe Ala Leu Leu Asp Arg Cys Gln
130 135 140

Gln Asp Ala Ile Phe Gln Val Val Trp Ser Glu Ile Phe Val Leu Arg
145 150 155 160

Ala Ser His Trp Ser Leu Asp Ile Ser Ala Met Ile Asp Gly Cys Gly
165 170 175

Asp Glu Gln Leu Lys Arg Leu Ile Cys Glu Ala His Gln Leu Arg Ala
180 185 190

Asp Val Leu Glu Leu Asn Phe Met Glu Ser Leu Ile Leu Cys Arg Lys
195 200 205

Glu Leu Ala Ile Asn Ala Glu Tyr Ala Val Ile Leu Gly Ser His Ser
210 215 220

Lys Ala Ala Leu Ile Ser Leu Ala Arg Tyr Thr Leu Gln Gln Ser Asn
225 230 235 240

Tyr Leu Arg Phe Gly Gln Leu Leu Leu Gly Leu Arg Gln Leu Cys Leu
245 250 255

Arg Arg Phe Asp Cys Ala Leu Ser Cys Met Phe Arg Ser Val Val Arg
260 265 270

Asp Ile Leu Lys Thr Leu
275

<210> 9
<211> 1626
<212> DNA
<213> Drosophila melanogaster
<400> 9
atgcacggac aggcgcctcc acctacatca acgggcgtgg ccccgcccac acagccaccg 60

ccccctcatc ccgccgcccc aaacgtgccc aatggtcgat tgctgagctg gaatcacagt 120

gccgctgcag ctgctgcggc ggcggcagcc caagcggcag ccaactccat gaaccactcg 180

tcggcggcgg agggttcatc gatgacccgg attaagggtc agaacctggg cctcatctgc 240

gtggtgtgcg gcgacaccag ctcgggaaag cactacggaa tcctagcctg caatggctgc 300

tccggattct tcaaacgcag cgcgggaacg ggacgctgtg tggtggacaa agctcatcgg 360

aatcaatgcc aggcctgcag gctcaagaag tgccttcaaa tgggaatgaa caaggacgcc 420

gtgcaaaatg agcgacagcc ccgcaacacg gccactatac ggccggagac tctgcgggag 480

atggagcacg gacgggcgct gcgagaggcc gccgtggccg tcggggtttt cgggccaccc 540

gtgctactgt ctccgccctg ctacggctcc ggactcctgc cgccgccctc actgggcagt 600

ttgccgacgg gacgattgct gcaccacaac cacctcacct cctcaatgca attggctgcg 660

aaccacatgg gcgccggcag ctttcccatg ttcaacgcag ccggcgtgca ccactcgccc 720

aaggaaaagg cctacggcat ggagatggcc accagtggca atgtctccca ctcgaccaac 780

tcgagcagca accactccat cgacccgagc tcaccggcgc cggaaaacgc caaggagata 840

aacatcgccg gcggcagtgt ctcctccgtc agctcttcca gtcccaccat ggaaaatgac 900

aatgatgacg actccataga tgtaaccaac gacaacgagg agccgcatgc agtcagcaga 960

tcggattcga gtttcattat gccgcagttc atgtcgccca atctgtacac ccatcaacac 1020

gaaacagttt acgagacaag tgcccggctg ctcttcatgg ccgtcaagtg ggccaagaac 1080

ctgcccagct ttgcaagact ttcctttcgg gatcaggtga ttttgctgga ggagtcctgg 1140

tcggagctgt tcctgctgaa cgcaatccaa tggtgcattc ccctggatcc caccggctgc 1200

gccctcttct cggtggcgga gcactgcaat aatctagaga acaatgccaa tggcgacact 1260

tgcataacaa aggaggagct ggcggcggat gtgcgaacgc tccacgagat cttctgcaaa 1320

tacaaggcgg tgctggtgga ccccgctgaa ttcgcgtgcc tcaaggcgat agttctcttc 1380

cggccggaaa cgcgcggact taaagatccg gcgcagatag agaatcttca ggatcaggcg 1440

cacgtaatgc tgtcgcagca cacaaagacg cagttcaccg cccagatagc cagattcgga 1500

cgactccttc tcatgctgcc gttgctgcgc atgatcagct cccacaagat tgagtccatc 1560

tattttcagc gcactattgg gaacacgccc atggaaaagg tgctctgtga catgtataag 1620

aactag 1626

<210> 10
<211> 541
<212> PRT
<213> Drosophila melanogaster
<400> 10

Met His Gly Gln Ala Pro Pro Pro Thr Ser Thr Gly Val Ala Pro Pro
1 5 10 15

Thr Gln Pro Pro Pro Pro His Pro Ala Ala Pro Asn Val Pro Asn Gly
20 25 30

```
Arg Leu Leu Ser Trp Asn His Ser Ala Ala Ala Ala Ala Ala Ala Ala
        35                  40                  45

Ala Ala Gln Ala Ala Ala Asn Ser Met Asn His Ser Ser Ala Ala Glu
    50                  55                  60

Gly Ser Ser Met Thr Arg Ile Lys Gly Gln Asn Leu Gly Leu Ile Cys
65                  70                  75                  80

Val Val Cys Gly Asp Thr Ser Ser Gly Lys His Tyr Gly Ile Leu Ala
                85                  90                  95

Cys Asn Gly Cys Ser Gly Phe Phe Lys Arg Ser Ala Gly Thr Gly Arg
            100                 105                 110

Cys Val Val Asp Lys Ala His Arg Asn Gln Cys Gln Ala Cys Arg Leu
        115                 120                 125

Lys Lys Cys Leu Gln Met Gly Met Asn Lys Asp Ala Val Gln Asn Glu
    130                 135                 140

Arg Gln Pro Arg Asn Thr Ala Thr Ile Arg Pro Glu Thr Leu Arg Glu
145                 150                 155                 160

Met Glu His Gly Arg Ala Leu Arg Glu Ala Ala Val Ala Val Gly Val
                165                 170                 175

Phe Gly Pro Pro Val Leu Leu Ser Pro Pro Cys Tyr Gly Ser Gly Leu
            180                 185                 190

Leu Pro Pro Pro Ser Leu Gly Ser Leu Pro Thr Gly Arg Leu Leu His
        195                 200                 205

His Asn His Leu Thr Ser Ser Met Gln Leu Ala Ala Asn His Met Gly
    210                 215                 220

Ala Gly Ser Phe Pro Met Phe Asn Ala Ala Gly Val His His Ser Pro
225                 230                 235                 240

Lys Glu Lys Ala Tyr Gly Met Glu Met Ala Thr Ser Gly Asn Val Ser
                245                 250                 255

His Ser Thr Asn Ser Ser Ser Asn His Ser Ile Asp Pro Ser Ser Pro
            260                 265                 270

Ala Pro Glu Asn Ala Lys Glu Ile Asn Ile Ala Gly Gly Ser Val Ser
        275                 280                 285

Ser Val Ser Ser Ser Ser Pro Thr Met Glu Asn Asp Asn Asp Asp Asp
    290                 295                 300

Ser Ile Asp Val Thr Asn Asp Asn Glu Glu Pro His Ala Val Ser Arg
305                 310                 315                 320

Ser Asp Ser Ser Phe Ile Met Pro Gln Phe Met Ser Pro Asn Leu Tyr
                325                 330                 335

Thr His Gln His Glu Thr Val Tyr Glu Thr Ser Ala Arg Leu Leu Phe
            340                 345                 350
```

Met Ala Val Lys Trp Ala Lys Asn Leu Pro Ser Phe Ala Arg Leu Ser
355 360 365

Phe Arg Asp Gln Val Ile Leu Leu Glu Glu Ser Trp Ser Glu Leu Phe
370 375 380

Leu Leu Asn Ala Ile Gln Trp Cys Ile Pro Leu Asp Pro Thr Gly Cys
385 390 395 400

Ala Leu Phe Ser Val Ala Glu His Cys Asn Asn Leu Glu Asn Asn Ala
405 410 415

Asn Gly Asp Thr Cys Ile Thr Lys Glu Glu Leu Ala Ala Asp Val Arg
420 425 430

Thr Leu His Glu Ile Phe Cys Lys Tyr Lys Ala Val Leu Val Asp Pro
435 440 445

Ala Glu Phe Ala Cys Leu Lys Ala Ile Val Leu Phe Arg Pro Glu Thr
450 455 460

Arg Gly Leu Lys Asp Pro Ala Gln Ile Glu Asn Leu Gln Asp Gln Ala
465 470 475 480

His Val Met Leu Ser Gln His Thr Lys Thr Gln Phe Thr Ala Gln Ile
485 490 495

Ala Arg Phe Gly Arg Leu Leu Leu Met Leu Pro Leu Leu Arg Met Ile
500 505 510

Ser Ser His Lys Ile Glu Ser Ile Tyr Phe Gln Arg Thr Ile Gly Asn
515 520 525

Thr Pro Met Glu Lys Val Leu Cys Asp Met Tyr Lys Asn
530 535 540

<210> 11
<211> 1599
<212> DNA
<213> Drosophila melanogaster
<400> 11
atggcgaccg ggcgttctct gctctttcga gtgccttggt atgtgtgctt gtgtgtgtgc 60

gcagagagcg cagagccggg tgtttattgg agattgcgat tgcggcttgg cttacccaca 120

ctcgcagggc cgcacaccaa cacactaaca ctaacagcga ggacaagctc ctgccgcagc 180

atcaagaagg aacgaatcaa agcaagccaa caagcaaatg cgccaccaga gttgccacta 240

aaagtctccg ttgacgttaa catcatcatc gcggcacact cgcagcgccg tcggatcgga 300

ttggttcggt ttcatcagcg ggaatcagag gaccgtccac ttgccgtcgc ctctccacga 360

ttgcaaatta atatggagcc tactgcgatg aacccgaaaa aactccacag tccgcagcgg 420

cattgctaca ctccgccgcc ggcgccgatg cacggacagg cgcctccacc tacatcaacg 480

ggcgtggccc cgcccacaca gccaccgccc cctcatcccg ccgccccaaa cgtgcccaat 540

ggtcgattgc tgagctggaa tcacagtgcc gctgcagctg ctgcggcggc ggcagcccaa 600

gcggcagcca actccatgaa ccactcgtcg gcggcggagg gttcatcgat gacccggatt 660

aagggtcaga acctgggcct catctgcgtg gtgtgcggcg acaccagctc gggaaagcac 720

tacggaatcc tagcctgcaa tggctgctcc ggattcttca aacgcagcgt gcggcggaaa 780

ctcatttatc gctgccaggc gggaacggga cgctgtgtgg tggacaaagc tcatcggaat 840

caatgccagg cctgcaggct caagaagtgc cttcaaatgg gaatgaacaa ggacgacgac 900

tccatagatg taaccaacga caacgaggag ccgcatgcag tcagcagatc ggattcgagt 960

ttcattatgc cgcagttcat gtcgcccaat ctgtacaccc atcaacacga aacagtttac 1020

gagacaagtg cccggctgct cttcatggcc gtcaagtggg ccaagaacct gcccagcttt 1080

gcaagacttt cctttcggga tcaggtaatt ttgctggagg agtcctggtc ggagctgttc 1140

ctgctgaacg caatccaatg gtgcattccc ctggatccca ccggctgcgc cctcttctcg 1200

gtggcggagc actgcaataa tctagagaac aatgccaatg gcgacacttg cataacaaag 1260

gaggagctgg cggcggatgt gcgaacgctc cacgagatct tctgcaaata caaggcggtg 1320

ctggtggacc ccgctgaatt cgcgtgcctc aaggcgatag ttctcttccg gccggaaacg 1380

cgcggactta aagatccggc gcagatagag aatcttcagg atcaggcgca ccacacaaag 1440

acgcagttca ccgcccagat agccagattc ggacgactcc ttctcatgct gccgttgctg 1500

cgcatgatca gctcccacaa gattgagtcc atctattttc agcgcactat tgggaacacg 1560

cccatggaaa aggtgctctg tgacatgtat aagaactag 1599

<210> 12
<211> 532
<212> PRT
<213> Drosophila melanogaster
<400> 12

Met Ala Thr Gly Arg Ser Leu Leu Phe Arg Val Pro Trp Tyr Val Cys
1 5 10 15

Leu Cys Val Cys Ala Glu Ser Ala Glu Pro Gly Val Tyr Trp Arg Leu
20 25 30

Arg Leu Arg Leu Gly Leu Pro Thr Leu Ala Gly Pro His Thr Asn Thr
35 40 45

Leu Thr Leu Thr Ala Arg Thr Ser Ser Cys Arg Ser Ile Lys Lys Glu
50 55 60

Arg Ile Lys Ala Ser Gln Gln Ala Asn Ala Pro Pro Glu Leu Pro Leu
65 70 75 80

Lys Val Ser Val Asp Val Asn Ile Ile Ile Ala Ala His Ser Gln Arg
85 90 95

```
Arg Arg Ile Gly Leu Val Arg Phe His Gln Arg Glu Ser Glu Asp Arg
            100                 105                 110

Pro Leu Ala Val Ala Ser Pro Arg Leu Gln Ile Asn Met Glu Pro Thr
        115                 120                 125

Ala Met Asn Pro Lys Lys Leu His Ser Pro Gln Arg His Cys Tyr Thr
    130                 135                 140

Pro Pro Pro Ala Pro Met His Gly Gln Ala Pro Pro Pro Thr Ser Thr
145                 150                 155                 160

Gly Val Ala Pro Pro Thr Gln Pro Pro Pro Pro His Pro Ala Ala Pro
                165                 170                 175

Asn Val Pro Asn Gly Arg Leu Leu Ser Trp Asn His Ser Ala Ala Ala
            180                 185                 190

Ala Ala Ala Ala Ala Ala Ala Gln Ala Ala Ala Asn Ser Met Asn His
        195                 200                 205

Ser Ser Ala Ala Glu Gly Ser Ser Met Thr Arg Ile Lys Gly Gln Asn
    210                 215                 220

Leu Gly Leu Ile Cys Val Val Cys Gly Asp Thr Ser Ser Gly Lys His
225                 230                 235                 240

Tyr Gly Ile Leu Ala Cys Asn Gly Cys Ser Gly Phe Phe Lys Arg Ser
                245                 250                 255

Val Arg Arg Lys Leu Ile Tyr Arg Cys Gln Ala Gly Thr Gly Arg Cys
            260                 265                 270

Val Val Asp Lys Ala His Arg Asn Gln Cys Gln Ala Cys Arg Leu Lys
        275                 280                 285

Lys Cys Leu Gln Met Gly Met Asn Lys Asp Asp Asp Ser Ile Asp Val
    290                 295                 300

Thr Asn Asp Asn Glu Glu Pro His Ala Val Ser Arg Ser Asp Ser Ser
305                 310                 315                 320

Phe Ile Met Pro Gln Phe Met Ser Pro Asn Leu Tyr Thr His Gln His
                325                 330                 335

Glu Thr Val Tyr Glu Thr Ser Ala Arg Leu Leu Phe Met Ala Val Lys
            340                 345                 350

Trp Ala Lys Asn Leu Pro Ser Phe Ala Arg Leu Ser Phe Arg Asp Gln
        355                 360                 365

Val Ile Leu Leu Glu Glu Ser Trp Ser Glu Leu Phe Leu Leu Asn Ala
    370                 375                 380

Ile Gln Trp Cys Ile Pro Leu Asp Pro Thr Gly Cys Ala Leu Phe Ser
385                 390                 395                 400

Val Ala Glu His Cys Asn Asn Leu Glu Asn Asn Ala Asn Gly Asp Thr
                405                 410                 415
```

Cys Ile Thr Lys Glu Glu Leu Ala Ala Asp Val Arg Thr Leu His Glu
420 425 430

Ile Phe Cys Lys Tyr Lys Ala Val Leu Val Asp Pro Ala Glu Phe Ala
435 440 445

Cys Leu Lys Ala Ile Val Leu Phe Arg Pro Glu Thr Arg Gly Leu Lys
450 455 460

Asp Pro Ala Gln Ile Glu Asn Leu Gln Asp Gln Ala His His Thr Lys
465 470 475 480

Thr Gln Phe Thr Ala Gln Ile Ala Arg Phe Gly Arg Leu Leu Leu Met
485 490 495

Leu Pro Leu Leu Arg Met Ile Ser Ser His Lys Ile Glu Ser Ile Tyr
500 505 510

Phe Gln Arg Thr Ile Gly Asn Thr Pro Met Glu Lys Val Leu Cys Asp
515 520 525

Met Tyr Lys Asn
530

<210> 13
<211> 4563
<212> DNA
<213> Drosophila melanogaster
<400> 13
atgacactga gccgtggccc gtacagcgag ctcgataaaa tgagcctttt tcaagacctc 60

aaactcaaac ggcgcaaaat cgattcgcga tgcagcagtg acggcgagtc catagcggac 120

acgtccacct cgtcgccgga cctgctggcg cccatgtcgc cgaagctctg cgacagcggc 180

tcggcggggg cgtcgctggg ggcatcgctg cccctgccgc tggccctgcc cctgccaatg 240

gccctgccac tgcccatgtc gctgcccctg cccctcacgg cggcatcttc ggcggtcacc 300

gtttcgctgg cagcggtcgt ggccgcggtg gccgagacgg gtggcgcggg cgcgggagga 360

gctgggacag cagtaacagc gtcgggagca ggaccatgcg tctccacgtc gtctacgacg 420

gcagcggcag ccacatcctc gacctcctcg ctctcgtcct cctcctcttc gtcatcctcc 480

acgtcctcca gcacttcctc cgcctcgccg acagctggag cctcctccac ggccacctgc 540

cccgccagca gcagcagcag cagtggaaac ggaagtgggg gcaaaagtgg tagcatcaag 600

caggagcaca cggagataca ctcgtcgagc agtgcgattt cggcggccgc cgcctcaacg 660

gtgatgtcac cgccgcccgc tgaggcgacg agatccagtc cagccacgcc cgagggaggc 720

ggaccagctg gcgacggaag tggagcaacg ggaggcggaa acacgagcgg cggatcaacg 780

gctggagtgg ccattaatga acaccaaaac aatggcaatg gcagcggcgg gagcagtcga 840

gcctctcccg attcgctgga agagaagccc tctaccacaa cgaccacagg tcgtccaacg 900

ctcacgccca cgaatggggt gctgtcctcc gcctcggcgg gcacggggat ttccacagga 960

```
agcagcgcca agctgagcga ggctggtatg agtgtgatac ggtccgtgaa ggaggagcgc   1020
ttgctcaacg tatccagcaa gatgctggtg ttccatcagc agcgggagca agagaccaaa   1080
gcagtggcgg ctgcagcagc agcagcagcg gcgggccatg tgacggttct agtgacgcca   1140
tcgcgcatca aatcggagcc accgccgccg gcttcaccct cctctacatc cagcacacaa   1200
agggaaaggg aacgggaacg cgatcgagag agggatcgcg aaagggaacg cgagcgggac   1260
cgggaccggg aacgggaacg ggaacagtcc atcagctcct cgcagcagca cctaagtcgg   1320
gtctccgcca gtccacccac tcagctgtcc cacggcagcc tgggacccaa cattgtgcag   1380
acgcaccatc ttcaccagca actcacacag ccgctgacgc tgcgcaagag cagcccgccc   1440
acagagcacc tgctcagtca gtccatgcaa catctcacac agcagcaggc gatccacctg   1500
catcacctac ttggccagca gcagcagcag cagcaggcgt cgcatcccca gcagcaacag   1560
cagcagcaac actcgcccca ctccctggtg cgggtgaaaa aggaaccgaa tgttggtcag   1620
cggcacttat cgccgcatca ccaacaacag tcgccactcc tgcagcacca ccaacagcag   1680
cagcagcagc aacaacaaca gcaacagcat ctgcatcagc aacagcaaca gcagcagcat   1740
caccagcagc agccccaggc actggccctg atgcatccgg cttccctggc gctaaggaac   1800
agcaatcggg atgcggccat tctgtttcgg gtgaagagcg aagtgcacca gcaggtggcc   1860
gccgggctgc cgcatctgat gcagtccgct ggtggggcag cggccgccgc cgcagcagct   1920
gtggccgctc agcgaatggt atgcttcagc aatgccagga tcaatggcgt taagccggag   1980
gtgattggag gaccgctggg caacctgcgg cccgtgggcg tcggtggcgg aaacggaagt   2040
ggctccgtgc agtgccctc gccgcatcca tcctcctcgt cgtcatcctc gcagctgtcg   2100
ccgcagacgc cctcccagac gccgccccga ggcacgccca ccgtcataat gggcgagagc   2160
tgcggggtgc gcaccatggt ctggggctac gagcctccgc caccctcggc gggccagtcc   2220
cacggccagc acccgcaaca gcaacagcag tcgccccacc accagccgca acaacaacag   2280
cagcagcaac aacagcagtc gcagcagcaa cagcaacagc agcagcaaca gtcgctgggc   2340
cagcagcagc actgcctctc ctcgccgtcg gcgggatcgc tgacgccctc ctcttcgtcc   2400
ggcggtggtt cggtatctgg cggcggagtg ggcggaccac tcacaccctc ctcggtggcg   2460
ccgcagaata acgaggaggc cgcccaactc ctgcttctcc ctgggacaga cacgcatcca   2520
ggacatgaga tcacgggcca cacccccttc cgcacaccgc acgcccttaa tatggagcgg   2580
ctgtgggcgg gagactactc gcaattgccg cccggccagc tgcaggctct gaatctcagt   2640
gcccaacagc agcagtgggg cagcagcaac tccacgggtc ttggtggcgt aggcggcggc   2700
atgggcggac gcaacctgga ggcgccgcac gagccgaccg acgaggacga acagccgctc   2760
```

gtttgcatga tctgcgagga caaggccacc ggcctgcact acggcatcat cacctgcgag 2820

gggtgcaagg gcttcttcaa gcggacggtg cagaaccgac gagtctacac ctgcgtggcg 2880

gacggcacct gcgagataac caaagcacag cgcaaccgtt gtcagtattg tcgatttaag 2940

aagtgcatcg agcagggcat ggtgctgcaa gccgttcgcg aggatcgcat gccgggcggt 3000

cgcaacagtg gcgccgtcta caatttgtac aaggtgaagt acaagaagca caagaagacc 3060

aatcagaagc agcagcagca ggccgcccag cagcagcagc agcaggcggc ggcgcagcag 3120

cagcaccagc aacagcagca gcatcaacag caccagcaac atcagcaaca gcagttgcac 3180

tcgccgctcc accatcacca ccaccagggc caccagtcgc accacgcgca gcagcagcac 3240

cacccacagc tgtcgccgca ccacctgctg tcgccgcagc agcagcaact tgccgccgcg 3300

gtggcagcag ctgcgcagca ccaacagcaa cagcaacaac agcagcaaca gcagcagcag 3360

gccaagctga tgggcggcgt ggtggacatg aagcccatgt tcctcggccc cgctttgaag 3420

ccggagttgc tgcaagcacc ccccatgcac agtccggccc agcaacaaca acagcagcag 3480

cagcagcagc agcaacagca ggcctcgccg catctctcgc ttagctcacc gcaccagcag 3540

cagcagcagc agcagggaca gcaccaaaac caccaccagc aacaaggtgg gggtggcgga 3600

ggagctggtg gaggagctca actgccgccg cacctggtga acggaacgat actgaagacg 3660

gccctaacca atcccagcga gattgtacat ctgcgccacc gcctcgactc ggcggtcagt 3720

tcgtccaagg accgacagat ctcgtacgag cacgccttag gcatgatcca gacactgatc 3780

gactgcgacg cgatggagga catagccaca ctgccgcact tcagcgagtt ccttgaggac 3840

aagtcggaga ttagcgagaa actgtgcaac atcggcgatt ccatagtcca caagctggtg 3900

tcgtggacaa aaaagttgcc cttctacctg gagatcccgg tggagataca taccaaacta 3960

ctgacggaca agtggcacga gatccttatc ctgaccacgg ccgcctacca ggcgttgcat 4020

ggcaagcggc gtggcgaggg aggaggcagc aggcatggtt cgccggcgtc aacgccactg 4080

agcacgccca ctggtacgcc gttgagcaca ccgataccct cgcccgccca gccactgcac 4140

aaggacgacc cggagtttgt cagcgaggtg aactcgcacc tgagcacact gcaaacctgc 4200

ttgaccacgc taatgggcca gccgatagcg atggagcagc tgaagctgga cgtcgggcac 4260

atggtggaca agatgaccca gatcaccatc atgttccggc gaatcaagct caagatggag 4320

gagtacgtct gcctgaaggt ttacatactg ctaaacaaag cagaagtgga actggagagc 4380

atccaggagc ggtacgtcca ggtgctgcgc tcctacctgc aaaactcctc gccgcagaat 4440

ccgcaggcga ggctcagtga actgctctcc cacataccag agatccaggc tgcggctagc 4500

ctgctgctcg agagcaagat gttctatgtg cccttcgtgc tcaactcggc gagcataagg 4560

tag 4563

<210> 14
<211> 1520
<212> PRT
<213> Drosophila melanogaster
<400> 14

Met Thr Leu Ser Arg Gly Pro Tyr Ser Glu Leu Asp Lys Met Ser Leu
1 5 10 15

Phe Gln Asp Leu Lys Leu Lys Arg Arg Lys Ile Asp Ser Arg Cys Ser
20 25 30

Ser Asp Gly Glu Ser Ile Ala Asp Thr Ser Thr Ser Ser Pro Asp Leu
35 40 45

Leu Ala Pro Met Ser Pro Lys Leu Cys Asp Ser Gly Ser Ala Gly Ala
50 55 60

Ser Leu Gly Ala Ser Leu Pro Leu Pro Leu Ala Leu Pro Leu Pro Met
65 70 75 80

Ala Leu Pro Leu Pro Met Ser Leu Pro Leu Pro Leu Thr Ala Ala Ser
85 90 95

Ser Ala Val Thr Val Ser Leu Ala Ala Val Val Ala Ala Val Ala Glu
100 105 110

Thr Gly Gly Ala Gly Ala Gly Gly Ala Gly Thr Ala Val Thr Ala Ser
115 120 125

Gly Ala Gly Pro Cys Val Ser Thr Ser Ser Thr Thr Ala Ala Ala Ala
130 135 140

Thr Ser Ser Thr Ser Ser Leu Ser Ser Ser Ser Ser Ser Ser Ser Ser
145 150 155 160

Thr Ser Ser Ser Thr Ser Ser Ala Ser Pro Thr Ala Gly Ala Ser Ser
165 170 175

Thr Ala Thr Cys Pro Ala Ser Ser Ser Ser Ser Ser Gly Asn Gly Ser
180 185 190

Gly Gly Lys Ser Gly Ser Ile Lys Gln Glu His Thr Glu Ile His Ser
195 200 205

Ser Ser Ser Ala Ile Ser Ala Ala Ala Ala Ser Thr Val Met Ser Pro
210 215 220

Pro Pro Ala Glu Ala Thr Arg Ser Ser Pro Ala Thr Pro Glu Gly Gly
225 230 235 240

Gly Pro Ala Gly Asp Gly Ser Gly Ala Thr Gly Gly Gly Asn Thr Ser
245 250 255

Gly Gly Ser Thr Ala Gly Val Ala Ile Asn Glu His Gln Asn Asn Gly
260 265 270

Asn Gly Ser Gly Gly Ser Ser Arg Ala Ser Pro Asp Ser Leu Glu Glu
275 280 285

```
Lys Pro Ser Thr Thr Thr Thr Thr Gly Arg Pro Thr Leu Thr Pro Thr
    290                 295                 300

Asn Gly Val Leu Ser Ser Ala Ser Ala Gly Thr Gly Ile Ser Thr Gly
305                 310                 315                 320

Ser Ser Ala Lys Leu Ser Glu Ala Gly Met Ser Val Ile Arg Ser Val
                325                 330                 335

Lys Glu Glu Arg Leu Leu Asn Val Ser Ser Lys Met Leu Val Phe His
            340                 345                 350

Gln Gln Arg Glu Gln Glu Thr Lys Ala Val Ala Ala Ala Ala Ala Ala
        355                 360                 365

Ala Ala Ala Gly His Val Thr Val Leu Val Thr Pro Ser Arg Ile Lys
    370                 375                 380

Ser Glu Pro Pro Pro Pro Ala Ser Pro Ser Ser Thr Ser Ser Thr Gln
385                 390                 395                 400

Arg Glu Arg Glu Arg Glu Arg Asp Arg Glu Arg Asp Arg Glu Arg Glu
                405                 410                 415

Arg Glu Arg Asp Arg Asp Arg Glu Arg Glu Arg Glu Gln Ser Ile Ser
            420                 425                 430

Ser Ser Gln Gln His Leu Ser Arg Val Ser Ala Ser Pro Pro Thr Gln
        435                 440                 445

Leu Ser His Gly Ser Leu Gly Pro Asn Ile Val Gln Thr His His Leu
    450                 455                 460

His Gln Gln Leu Thr Gln Pro Leu Thr Leu Arg Lys Ser Ser Pro Pro
465                 470                 475                 480

Thr Glu His Leu Leu Ser Gln Ser Met Gln His Leu Thr Gln Gln Gln
                485                 490                 495

Ala Ile His Leu His His Leu Leu Gly Gln Gln Gln Gln Gln Gln Gln
            500                 505                 510

Ala Ser His Pro Gln Gln Gln Gln Gln Gln Gln His Ser Pro His Ser
        515                 520                 525

Leu Val Arg Val Lys Lys Glu Pro Asn Val Gly Gln Arg His Leu Ser
    530                 535                 540

Pro His His Gln Gln Gln Ser Pro Leu Leu Gln His His Gln Gln Gln
545                 550                 555                 560

Gln Gln Gln Gln Gln Gln Gln Gln Gln His Leu His Gln Gln Gln Gln
                565                 570                 575

Gln Gln Gln His His Gln Gln Gln Pro Gln Ala Leu Ala Leu Met His
            580                 585                 590

Pro Ala Ser Leu Ala Leu Arg Asn Ser Asn Arg Asp Ala Ala Ile Leu
        595                 600                 605
```

```
Phe Arg Val Lys Ser Glu Val His Gln Gln Val Ala Ala Gly Leu Pro
    610                 615                 620

His Leu Met Gln Ser Ala Gly Gly Ala Ala Ala Ala Ala Ala Ala Ala
625                 630                 635                 640

Val Ala Ala Gln Arg Met Val Cys Phe Ser Asn Ala Arg Ile Asn Gly
                645                 650                 655

Val Lys Pro Glu Val Ile Gly Gly Pro Leu Gly Asn Leu Arg Pro Val
            660                 665                 670

Gly Val Gly Gly Gly Asn Gly Ser Gly Ser Val Gln Cys Pro Ser Pro
        675                 680                 685

His Pro Ser Ser Ser Ser Ser Ser Ser Gln Leu Ser Pro Gln Thr Pro
    690                 695                 700

Ser Gln Thr Pro Pro Arg Gly Thr Pro Thr Val Ile Met Gly Glu Ser
705                 710                 715                 720

Cys Gly Val Arg Thr Met Val Trp Gly Tyr Glu Pro Pro Pro Pro Ser
                725                 730                 735

Ala Gly Gln Ser His Gly Gln His Pro Gln Gln Gln Gln Gln Ser Pro
            740                 745                 750

His His Gln Pro Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Ser Gln
        755                 760                 765

Gln Gln Gln Gln Gln Gln Gln Gln Gln Ser Leu Gly Gln Gln Gln His
    770                 775                 780

Cys Leu Ser Ser Pro Ser Ala Gly Ser Leu Thr Pro Ser Ser Ser Ser
785                 790                 795                 800

Gly Gly Gly Ser Val Ser Gly Gly Gly Val Gly Gly Pro Leu Thr Pro
                805                 810                 815

Ser Ser Val Ala Pro Gln Asn Asn Glu Glu Ala Ala Gln Leu Leu Leu
            820                 825                 830

Leu Pro Gly Thr Asp Thr His Pro Gly His Glu Ile Thr Gly His Thr
        835                 840                 845

Pro Phe Arg Thr Pro His Ala Leu Asn Met Glu Arg Leu Trp Ala Gly
    850                 855                 860

Asp Tyr Ser Gln Leu Pro Pro Gly Gln Leu Gln Ala Leu Asn Leu Ser
865                 870                 875                 880

Ala Gln Gln Gln Gln Trp Gly Ser Ser Asn Ser Thr Gly Leu Gly Gly
                885                 890                 895

Val Gly Gly Gly Met Gly Gly Arg Asn Leu Glu Ala Pro His Glu Pro
            900                 905                 910

Thr Asp Glu Asp Glu Gln Pro Leu Val Cys Met Ile Cys Glu Asp Lys
        915                 920                 925
```

```
Ala Thr Gly Leu His Tyr Gly Ile Ile Thr Cys Glu Gly Cys Lys Gly
    930                 935                 940

Phe Phe Lys Arg Thr Val Gln Asn Arg Arg Val Tyr Thr Cys Val Ala
945                 950                 955                 960

Asp Gly Thr Cys Glu Ile Thr Lys Ala Gln Arg Asn Arg Cys Gln Tyr
                965                 970                 975

Cys Arg Phe Lys Lys Cys Ile Glu Gln Gly Met Val Leu Gln Ala Val
            980                 985                 990

Arg Glu Asp Arg Met Pro Gly Gly  Arg Asn Ser Gly Ala  Val Tyr Asn
        995                 1000                1005

Leu Tyr  Lys Val Lys Tyr Lys  Lys His Lys Lys Thr  Asn Gln Lys
    1010                1015                1020

Gln Gln  Gln Gln Ala Ala Gln  Gln Gln Gln Gln Gln  Ala Ala Ala
    1025                1030                1035

Gln Gln  Gln His Gln Gln Gln  Gln Gln His Gln Gln  His Gln Gln
    1040                1045                1050

His Gln  Gln Gln Gln Leu His  Ser Pro Leu His His  His His His
    1055                1060                1065

Gln Gly  His Gln Ser His His  Ala Gln Gln Gln His  His Pro Gln
    1070                1075                1080

Leu Ser  Pro His His Leu Leu  Ser Pro Gln Gln Gln  Gln Leu Ala
    1085                1090                1095

Ala Ala  Val Ala Ala Ala Ala  Gln His Gln Gln Gln  Gln Gln Gln
    1100                1105                1110

Gln Gln  Gln Gln Gln Gln Gln  Ala Lys Leu Met Gly  Gly Val Val
    1115                1120                1125

Asp Met  Lys Pro Met Phe Leu  Gly Pro Ala Leu Lys  Pro Glu Leu
    1130                1135                1140

Leu Gln  Ala Pro Pro Met His  Ser Pro Ala Gln Gln  Gln Gln Gln
    1145                1150                1155

Gln Gln  Gln Gln Gln Gln Gln  Gln Gln Ala Ser Pro  His Leu Ser
    1160                1165                1170

Leu Ser  Ser Pro His Gln Gln  Gln Gln Gln Gln Gln  Gly Gln His
    1175                1180                1185

Gln Asn  His His Gln Gln Gln  Gly Gly Gly Gly Gly  Gly Ala Gly
    1190                1195                1200

Gly Gly  Ala Gln Leu Pro Pro  His Leu Val Asn Gly  Thr Ile Leu
    1205                1210                1215

Lys Thr  Ala Leu Thr Asn Pro  Ser Glu Ile Val His  Leu Arg His
    1220                1225                1230
```

Arg Leu Asp Ser Ala Val Ser Ser Ser Lys Asp Arg Gln Ile Ser
1235 1240 1245

Tyr Glu His Ala Leu Gly Met Ile Gln Thr Leu Ile Asp Cys Asp
1250 1255 1260

Ala Met Glu Asp Ile Ala Thr Leu Pro His Phe Ser Glu Phe Leu
1265 1270 1275

Glu Asp Lys Ser Glu Ile Ser Glu Lys Leu Cys Asn Ile Gly Asp
1280 1285 1290

Ser Ile Val His Lys Leu Val Ser Trp Thr Lys Lys Leu Pro Phe
1295 1300 1305

Tyr Leu Glu Ile Pro Val Glu Ile His Thr Lys Leu Leu Thr Asp
1310 1315 1320

Lys Trp His Glu Ile Leu Ile Leu Thr Thr Ala Ala Tyr Gln Ala
1325 1330 1335

Leu His Gly Lys Arg Arg Gly Glu Gly Gly Gly Ser Arg His Gly
1340 1345 1350

Ser Pro Ala Ser Thr Pro Leu Ser Thr Pro Thr Gly Thr Pro Leu
1355 1360 1365

Ser Thr Pro Ile Pro Ser Pro Ala Gln Pro Leu His Lys Asp Asp
1370 1375 1380

Pro Glu Phe Val Ser Glu Val Asn Ser His Leu Ser Thr Leu Gln
1385 1390 1395

Thr Cys Leu Thr Thr Leu Met Gly Gln Pro Ile Ala Met Glu Gln
1400 1405 1410

Leu Lys Leu Asp Val Gly His Met Val Asp Lys Met Thr Gln Ile
1415 1420 1425

Thr Ile Met Phe Arg Arg Ile Lys Leu Lys Met Glu Glu Tyr Val
1430 1435 1440

Cys Leu Lys Val Tyr Ile Leu Leu Asn Lys Ala Glu Val Glu Leu
1445 1450 1455

Glu Ser Ile Gln Glu Arg Tyr Val Gln Val Leu Arg Ser Tyr Leu
1460 1465 1470

Gln Asn Ser Ser Pro Gln Asn Pro Gln Ala Arg Leu Ser Glu Leu
1475 1480 1485

Leu Ser His Ile Pro Glu Ile Gln Ala Ala Ala Ser Leu Leu Leu
1490 1495 1500

Glu Ser Lys Met Phe Tyr Val Pro Phe Val Leu Asn Ser Ala Ser
1505 1510 1515

Ile Arg
1520

<210> 15
<211> 693
<212> PRT
<213> Drosophila melanogaster
<400> 15

Met Gly Thr Ala Gly Asp Arg Leu Leu Asp Ile Pro Cys Lys Val Cys
1 5 10 15

Gly Asp Arg Ser Ser Gly Lys His Tyr Gly Ile Tyr Ser Cys Asp Gly
20 25 30

Cys Ser Gly Phe Phe Lys Arg Ser Ile His Arg Asn Arg Ile Tyr Thr
35 40 45

Cys Lys Ala Thr Gly Asp Leu Lys Gly Arg Cys Pro Val Asp Lys Thr
50 55 60

His Arg Asn Gln Cys Arg Ala Cys Arg Leu Ala Lys Cys Phe Gln Ser
65 70 75 80

Ala Met Asn Lys Asp Ala Val Gln His Glu Arg Gly Pro Arg Lys Pro
85 90 95

Lys Leu His Pro Gln Leu His His His His His His Ala Ala Ala Ala
100 105 110

Ala Ala Ala Ala His His Ala Ala Ala Ala His His His His His His
115 120 125

His His His Ala His Ala Ala Ala Ala His His Ala Ala Val Ala Ala
130 135 140

Ala Ala Ala Ser Gly Leu His His His His His Ala Met Pro Val Ser
145 150 155 160

Leu Val Thr Asn Val Ser Ala Ser Phe Asn Tyr Thr Gln His Ile Ser
165 170 175

Thr His Pro Pro Ala Pro Ala Ala Pro Pro Ser Gly Phe His Leu Thr
180 185 190

Ala Ser Gly Ala Gln Gln Gly Pro Ala Pro Pro Ala Gly His Leu His
195 200 205

His Gly Gly Ala Gly His Gln His Ala Thr Ala Phe His His Pro Gly
210 215 220

His Gly His Ala Leu Pro Ala Pro His Gly Gly Val Ile Ser Asn Pro
225 230 235 240

Gly Gly Asn Ser Ser Ala Ile Ser Gly Ser Gly Pro Gly Ser Thr Leu
245 250 255

Pro Phe Pro Ser His Leu Leu His His Asn Leu Ile Ala Glu Ala Ala
260 265 270

Ser Lys Leu Pro Gly Ile Thr Ala Thr Ala Val Ala Ala Val Val Ser
275 280 285

```
Ser Thr Ser Thr Pro Tyr Ala Ser Ala Ala Gln Ala Ser Ser Pro Ser
    290                 295                 300

Ser Asn Asn His Asn Tyr Ser Ser Pro Ser Pro Ser Asn Ser Ile Gln
305                 310                 315                 320

Ser Ile Ser Ser Ile Gly Ser Arg Ser Gly Gly Gly Glu Glu Gly Leu
                325                 330                 335

Ser Leu Gly Ser Glu Ser Pro Arg Val Asn Val Glu Thr Glu Thr Pro
            340                 345                 350

Ser Pro Ser Asn Ser Pro Pro Leu Ser Ala Gly Ser Ile Ser Pro Ala
        355                 360                 365

Pro Thr Leu Thr Thr Ser Ser Gly Ser Pro Gln His Arg Gln Met Ser
    370                 375                 380

Arg His Ser Leu Ser Glu Ala Thr Thr Pro Pro Ser His Ala Ser Leu
385                 390                 395                 400

Met Ile Cys Ala Ser Asn Asn Asn Asn Asn Asn Asn Asn Asn Asn Asn
                405                 410                 415

Asn Asn Asn Gly Glu His Lys Gln Ser Ser Tyr Thr Ser Gly Ser Pro
            420                 425                 430

Thr Pro Thr Thr Pro Thr Pro Pro Pro Pro Arg Ser Gly Val Gly Ser
        435                 440                 445

Thr Cys Asn Thr Ala Ser Ser Ser Ser Gly Phe Leu Glu Leu Leu Leu
    450                 455                 460

Ser Pro Asp Lys Cys Gln Glu Leu Ile Gln Tyr Gln Val Gln His Asn
465                 470                 475                 480

Thr Leu Leu Phe Pro Gln Gln Leu Leu Asp Ser Arg Leu Leu Ser Trp
                485                 490                 495

Glu Met Leu Gln Glu Thr Thr Ala Arg Leu Leu Phe Met Ala Val Arg
            500                 505                 510

Trp Val Lys Cys Leu Met Pro Phe Gln Thr Leu Ser Lys Asn Asp Gln
        515                 520                 525

His Leu Leu Leu Gln Glu Ser Trp Lys Glu Leu Phe Leu Leu Asn Leu
    530                 535                 540

Ala Gln Trp Thr Ile Pro Leu Asp Leu Thr Pro Ile Leu Glu Ser Pro
545                 550                 555                 560

Leu Ile Arg Glu Arg Val Leu Gln Asp Glu Ala Thr Gln Thr Glu Met
                565                 570                 575

Lys Thr Ile Gln Glu Ile Leu Cys Arg Phe Arg Gln Ile Thr Pro Asp
            580                 585                 590

Gly Ser Glu Val Gly Cys Met Lys Ala Ile Ala Leu Phe Ala Pro Glu
        595                 600                 605
```

Thr Ala Gly Leu Cys Asp Val Gln Pro Val Glu Met Leu Gln Asp Gln
610 615 620

Ala Gln Cys Ile Leu Ser Asp His Val Arg Leu Arg Tyr Pro Arg Gln
625 630 635 640

Ala Thr Arg Phe Gly Arg Leu Leu Leu Leu Leu Pro Ser Leu Arg Thr
645 650 655

Ile Arg Ala Ala Thr Ile Glu Ala Leu Phe Phe Lys Glu Thr Ile Gly
660 665 670

Asn Val Pro Ile Ala Arg Leu Leu Arg Asp Met Tyr Thr Met Glu Pro
675 680 685

Ala Gln Val Asp Lys
690

<210> 16
<211> 3049
<212> DNA
<213> Drosophila melanogaster
<400> 16
gtcagcccag gcgatccgca tttgcgtccg cagcaggttt ccgatttcag aactctgatt 60

ccagcggcag cgaatcgcgt cggcatctga acatttgaaa ataatctaaa attgcaagtg 120

actttgtgca ccggttacac taaaattgtt aacaaatcgc catatattct gaatttaaat 180

ttaaagtgcg cagtgcggaa tataaatcgg agcaaactgg atacgttagg gttcaaatac 240

ttccatcaac ggaaaatggg cacagcgggc gatcgcctgt tggacattcc ctgcaaggtg 300

tgtggcgatc gcagctccgg caagcactat ggaatctaca gctgcgatgg ctgctccggt 360

tttttcaagc ggagcattca tcgcaatcgg atttacacct gtaaggccac cggcgatctc 420

aagggtcgct gtccggtgga caagacccat cggaatcagt gtcgcgcctg tcgcctggcc 480

aagtgcttcc agtcggccat gaacaaggat gctgtgcagc acgagcgcgg tcctaggaaa 540

cccaagttgc acccgcaact gcatcatcat catcatcatg ctgctgccgc cgccgctgca 600

gcgcatcatg cagcagccgc ccatcaccat caccatcatc accaccacgc ccacgcagcg 660

gccgcccatc atgcggcagt ggctgcagcg gctgcctccg ggctgcatca ccaccaccac 720

gccatgcccg tctcgctggt gaccaatgtc tcggcctcgt tcaactatac gcagcacatc 780

tccacgcatc cgcctgctcc ggcggcgcca cccagtggct ttcacctgac ggccagtggc 840

gcccagcagg gaccagctcc accagctggc cacctgcacc atggtggagc cggacatcag 900

cacgccacgg ccttccacca tccgggacat ggacacgcgt tgcctgcccc acatggcggc 960

gtgatcagca atcccggcgg caactcgagc gcaatctccg gcagcggacc cggctccacg 1020

ctgcccttcc cctcgcacct gctgcaccac aatctgatag cggaggcggc cagcaagctg 1080

ccgggcatca ctgccacagc cgttgcggcg gtggtgtcct ccactagcac gccctacgcc 1140

```
tcggcggccc aggcgtcgtc gcctagtagc aacaaccaca actactcctc gccctcgccc   1200
agcaactcca tccagtccat ctcgagcatt ggatcgcgca gcggtggtgg cgaggagggc   1260
ctcagcctgg gcagcgagag tccgcgcgtc aatgtggaaa ctgagacacc ttcgccctca   1320
aactcgccgc ccctgagtgc tggtagcatt tcgccagcgc ccacgttgac cacctcgtcg   1380
ggatcgccgc agcaccgcca gatgtcgcgg cacagcctca gtgaggcaac cacgccgccc   1440
agccacgcct ctctcatgat ttgcgccagc aacaacaaca acaacaacaa taataataac   1500
aacaacaata atggagagca caagcagtcg agctacacat ccggatcacc gacacccaca   1560
acgcccacgc cgccaccgcc gcgttctggt gtaggttcca cctgcaacac ggccagcagc   1620
tccagcggct tcctggagct gctgctcagt ccggacaagt gccaggagct catccagtac   1680
caggtgcagc acaacacgct gctcttcccg cagcagctgc tggactcgcg gctgctctcc   1740
tgggagatgc tgcaggagac gacggcgcga ctgctcttca tggcggtgcg ctgggtcaag   1800
tgcctcatgc cgttccagac gctctccaag aacgaccagc atctgctgct ccaggaatcc   1860
tggaaggagc tctttctgct taatctcgcc caatggacta taccgctgga tctaacgccc   1920
atactggaat caccgctcat ccgcgaacgg gtgctgcagg acgaggccac acagacggag   1980
atgaagacga tccaggagat cctgtgccgc ttccgccaga tcacacccga cggcagcgag   2040
gttggctgca tgaaggccat cgccctgttc gcacccgaaa ccgccggcct gtgcgacgtg   2100
cagccggtgg agatgttgca ggatcaggcg cagtgcatcc tctccgacca tgtgcgactg   2160
cgctacccgc gccaagcaac ccgcttcggc aggctgttgc tcctgctgcc ctcgctgcgc   2220
accatccggg cggccaccat cgaggcgctg ttcttcaagg agaccatcgg caatgtgccc   2280
attgctcgac tgctgcgcga catgtacacc atggaaccgg cacaggtgga caagtgaacc   2340
ggccacgcat ggccgtcgaa atgaaatcaa aatcgattcc ctagcaccta agcgccaccc   2400
atcggtcgtc gtcatatgcg aacttatttg tattccaatg ggacccgaat cgtattcaga   2460
ttcactgcgg caggaggcgg tccaaatgtg gggcggaagc tgcagatgct atggttcgca   2520
ggacgccatg taatggaggc gtatgtacta accgcgctcc tccattggcg atgcagtccg   2580
cgatgatggc gcactcccac acccacacca gtacccacac cttgatttat cgccggcaat   2640
gcgtcggagt ctccttactt tcgcttcgtt ttctaacatt tgtatcctta ttttatttca   2700
tctttttcca cggatttttc gttttgactg cctgggcggc actctttatt tatctttcat   2760
tcgacgtttt gtcgtcgctt ttctaaaaat tccccatgtt atttcaacct ggcaaggacc   2820
tcgcaatccc attcccgcgc ccttacttac aaatcacttc ccatcccaca tccagcaatt   2880
ccgtggtttg aattctttcg tgcattgact acgaaatacc ctttaatcag acaaataaag   2940
```

aatattagtt gtaattcttt tttctgcaat ccagctctaa aacgggtttc ttaatcgaaa 3000

tcgataaatg taaaaattat acatatcctt taccaacatt gtttgcctа 3049

<210> 17
<211> 1974
<212> DNA
<213> Heliothis virescens
<400> 17
ccgagcggca cgcgtgccgc ttgatgctaa ctattctcat tgcgatcgta acataccacg 60

cgctgtcaac acatgttcga acagtagtta tagtttcttt ggtgactgac acagacgctt 120

aacgcctgta aacttcttac accgctcgta tagaacaaaa tgtattggat ttatgtgctt 180

cctacaatag tgcttgaaca ggtgcagccc gcgagcggta gctagtgttg cggccagacc 240

gctagtgaac tcggtgactg tgtgaacgtt gttgcggttg aacgccgcct tagggatttg 300

taataaacag ccgcactgag cggccgagca tgacgatgga ccagcagaca ggcctcatgt 360

ccctcaatat gtccccgttt gatctgagcc caggccccga agggtctggt tcaggtgggg 420

gtccctcgag tgcatcgcaa cagtacgtgc cgcaaggcgc cgcgtatcaa tgccccctg 480

aacaacaatc tttcggctat gccaatctgg atgcttcata tctgtttcca acaggcgccg 540

gtggcgaacc aggcgcctac ttgcctgcag ccggcaccgt gtgtgatcag acagacacca 600

aggatgtgat cgaagaactg tgtcccgtct gtggagacaa agtcagcggt tatcactacg 660

ggctactgac gtgcgagtct tgtaagggct tcttcaaaag aaccgttcaa aacaagaagg 720

tgtacacgtg cgtcgctgaa cgagcctgcc acatagacaa aacacaacgg aaacgctgcc 780

cattttgccg cttccaaaag tgcctcgatg ttggcatgaa actcgaagct gtacgagctg 840

accgcatgcg cggtggtcgt aacaagttcg gccccatgta caaacgggat cgtgcccgca 900

aactgcagat gatgcgacag cgacaaatcg cagtgcagac attgcgcggc tcacttggtg 960

acagcggcct ggtgctcggc ttcgggtctc cttacgccac agttccagtc aagcaagaga 1020

tacagatccc tcaagtgtcg tctctcacgt cttcacccga gtcgtcgcct gggccggcct 1080

tactggccgc ccagccgcag ccaccacagc cgccgccgcc gccagctcac gacaagtggg 1140

aagcacattc gccgcactcc gcatctccgg acgccttcgc gtttgacgcg ccagccacag 1200

cagccgcaac tccatccagc accgccgaac ccacaagcac tgaaaccctg cgagtctcac 1260

ccatgatacg cgaatttgta caaactatcg atgatcgcga gtggcagaat tcgctatttg 1320

ggctcttgca gagccagact tacaatcagt gtgaggtcga tctctttgag ttaatgtgca 1380

aagtattgga ccaaaactta ttttcacaag tggattgggc gagaaacacc gtgttcttta 1440

agtatctaaa ggtcgatgat caaatgaagc tattgcaaca ctcgtggtcg gacatgctgg 1500

tgttagatca cctacaccag aggatgcaca atggcctgcc tgatgagact accctccaca 1560

acggacagaa gtttgacctc ctttgtctgg gactcctggg cgttccatcc ctggctgatc 1620

acttcaacga actccagaac aaattagcag agctcaaatt cgatgtaccg gactacatat 1680

gtgttaaatt cttgcttcta ttaaatcctg aggtgagagg catcgtgaat gtgaagtgcg 1740

tccgagatgg ttatcagaca gtccaagccg cacttctgga ctatacgctg tcctgttatc 1800

ctacgataca ggataaattc gacaaattag tgatggtagt acctgagatc catgccctag 1860

cagctcgagg agaagagcac ctataccagc ggcactgtgc aggccaggcg cccacacaaa 1920

cactcctaat ggaaatgcta cacgcgaagc gcaaatcttg aagtctcagt atgg 1974

<210> 18
<211> 543
<212> PRT
<213> Heliothis virescens

<400> 18

Met Thr Met Asp Gln Gln Thr Gly Leu Met Ser Leu Asn Met Ser Pro
1 5 10 15

Phe Asp Leu Ser Pro Gly Pro Glu Gly Ser Gly Ser Gly Gly Gly Pro
20 25 30

Ser Ser Ala Ser Gln Gln Tyr Val Pro Gln Gly Ala Ala Tyr Gln Cys
35 40 45

Pro Pro Glu Gln Gln Ser Phe Gly Tyr Ala Asn Leu Asp Ala Ser Tyr
50 55 60

Leu Phe Pro Thr Gly Ala Gly Gly Glu Pro Gly Ala Tyr Leu Pro Ala
65 70 75 80

Ala Gly Thr Val Cys Asp Gln Thr Asp Thr Lys Asp Val Ile Glu Glu
85 90 95

Leu Cys Pro Val Cys Gly Asp Lys Val Ser Gly Tyr His Tyr Gly Leu
100 105 110

Leu Thr Cys Glu Ser Cys Lys Gly Phe Phe Lys Arg Thr Val Gln Asn
115 120 125

Lys Lys Val Tyr Thr Cys Val Ala Glu Arg Ala Cys His Ile Asp Lys
130 135 140

Thr Gln Arg Lys Arg Cys Pro Phe Cys Arg Phe Gln Lys Cys Leu Asp
145 150 155 160

Val Gly Met Lys Leu Glu Ala Val Arg Ala Asp Arg Met Arg Gly Gly
165 170 175

Arg Asn Lys Phe Gly Pro Met Tyr Lys Arg Asp Arg Ala Arg Lys Leu
180 185 190

Gln Met Met Arg Gln Arg Gln Ile Ala Val Gln Thr Leu Arg Gly Ser
195 200 205

```
Leu Gly Asp Ser Gly Leu Val Leu Gly Phe Gly Ser Pro Tyr Ala Thr
    210                 215                 220

Val Pro Val Lys Gln Glu Ile Gln Ile Pro Gln Val Ser Ser Leu Thr
225                 230                 235                 240

Ser Ser Pro Glu Ser Ser Pro Gly Pro Ala Leu Leu Ala Ala Gln Pro
                245                 250                 255

Gln Pro Pro Gln Pro Pro Pro Pro Pro Ala His Asp Lys Trp Glu Ala
            260                 265                 270

His Ser Pro His Ser Ala Ser Pro Asp Ala Phe Ala Phe Asp Ala Pro
        275                 280                 285

Ala Thr Ala Ala Ala Thr Pro Ser Ser Thr Ala Glu Pro Thr Ser Thr
    290                 295                 300

Glu Thr Leu Arg Val Ser Pro Met Ile Arg Glu Phe Val Gln Thr Ile
305                 310                 315                 320

Asp Asp Arg Glu Trp Gln Asn Ser Leu Phe Gly Leu Leu Gln Ser Gln
                325                 330                 335

Thr Tyr Asn Gln Cys Glu Val Asp Leu Phe Glu Leu Met Cys Lys Val
            340                 345                 350

Leu Asp Gln Asn Leu Phe Ser Gln Val Asp Trp Ala Arg Asn Thr Val
        355                 360                 365

Phe Phe Lys Tyr Leu Lys Val Asp Asp Gln Met Lys Leu Leu Gln His
    370                 375                 380

Ser Trp Ser Asp Met Leu Val Leu Asp His Leu His Gln Arg Met His
385                 390                 395                 400

Asn Gly Leu Pro Asp Glu Thr Thr Leu His Asn Gly Gln Lys Phe Asp
                405                 410                 415

Leu Leu Cys Leu Gly Leu Leu Gly Val Pro Ser Leu Ala Asp His Phe
            420                 425                 430

Asn Glu Leu Gln Asn Lys Leu Ala Glu Leu Lys Phe Asp Val Pro Asp
        435                 440                 445

Tyr Ile Cys Val Lys Phe Leu Leu Leu Leu Asn Pro Glu Val Arg Gly
    450                 455                 460

Ile Val Asn Val Lys Cys Val Arg Asp Gly Tyr Gln Thr Val Gln Ala
465                 470                 475                 480

Ala Leu Leu Asp Tyr Thr Leu Ser Cys Tyr Pro Thr Ile Gln Asp Lys
                485                 490                 495

Phe Asp Lys Leu Val Met Val Val Pro Glu Ile His Ala Leu Ala Ala
            500                 505                 510

Arg Gly Glu Glu His Leu Tyr Gln Arg His Cys Ala Gly Gln Ala Pro
        515                 520                 525
```

Thr Gln Thr Leu Leu Met Glu Met Leu His Ala Lys Arg Lys Ser
530 535 540

<210> 19
<211> 2526
<212> DNA
<213> Heliothis virescens
<400> 19
cgccgccggc cgcacgcgcc ccatccgcgc ggctcatgcg tctcccgccg ctcccgctgc 60

ccgacatgac tgtgaccgag tgccagcgcc gactcctgga gcccagcgcc gcagagccgc 120

cgcccccgc gccccccact gattccgacg tgctgctcgg cagagttctc gcagagtttg 180

acggcactac ggtgctttgc cgcgtttgcg gtgataaagc tagtggattt cattacggcg 240

tgcattcctg tgaaggttgc aagggtttct tcaggcggtc tatccaacag aagatccaat 300

atagaccctg cacgaaaaac caacagtgct caatcctccg gatcaacaga aatcggtgcc 360

agtactgccg attgaaaaaa tgcatcgcag taggcatgag cagagatgcc gtgcggttcg 420

gacgtgttcc aaaacgcgag aaagcgcgta tccttgcagc aatgcagcaa tcttctacgt 480

cacgagcaaa tgagcaggct gcggccgcag agctcgatga tgctccgaga ttgctggcgc 540

gcgtggtgcg cgctcacctc gacacctgcg agttcacgcg ggatcgcgtc gccgctatgc 600

gtgcacgtgc tcgcgactgt cccatctact cgcaacccac tctggcttgc ccactaaacc 660

cggcgccaga gcttcagtct gagaaggaat tctctcagcg atttgcacac gttattcgtg 720

gtgtaataga cttcgctggc ctcatccccg gcttccagct tcttacacag gatgacaagt 780

ttacgctgct caaaagtgga ctatttgatg cgttatttgt gcggctaatc tgtatgtttg 840

acgcgccact caatagtata atttgtctaa atggccaagt catgaaacgg gactccatac 900

aaagtggtgc caatgcaagg ttccttgtgg attctacttt taaattcgcc gagcgcatga 960

actctatgaa cttaacagat gcagaaatcg ccctttttctg cgctatagtg ctcatcactc 1020

cggataggcc gggacttcgc aatgtagaac tcgtggaaag aatgcatgcg cgtctgaagg 1080

catgcctgca aactgtggtc gcccagaata gaccggaccg accaggtttc ctcagggaac 1140

ttatggatac ccttcccgac ctccgcacat taagcacact ccatactgaa aagctagtag 1200

ttttccgtac tgaacacaaa gaattactac ggcagcagat gtggaccgat gaagaaggtg 1260

tgatgtcctg gggtgattct ggtgccgatg agtcagcacg cagccccatt gggtctgtgt 1320

ccagcagcga gtccggcgaa gccgtcggtg actgcggcac gccactcttg gccgctacac 1380

tggccggacg ccggcgactc gactctcgtg gatccgtaga tgaagaagca ctaggcgtgg 1440

cacatctagc tcacaacgga ctcactgtaa caccagtccg cccgcctcca cgataccgca 1500

aacttgattc gccaactgat tctggcatcg agtcgggaaa tgagaagcac gagaggatag 1560

ttggccccgg atcaggatgc tcaagcccgc gatcttccct agaggagcac acggatgata 1620

ggcggccacc tccatcagcc gatgacatgc ctgtgctaaa acgcgtgcta gaagcgccac 1680

cgctgtacca caccacttca ttaatggacg aggcgtataa acctcacaag aagttccgtg 1740

caatgcgccg tgacaccggc gaagctgagg cccgtgtggt caggccagcg ccctccacgc 1800

agccaccgca gcacccgcac cccgcgagcc cggcgcaccc ggcgcactcg ccgcgcccac 1860

tgcgcgcctc gctgtcatcc acgcactccg tgctggcgaa gagcctgatg gaaggcccgc 1920

gtatgacccc ggagcagctg aagcgcaccg acatcatcca gcagtacatg cggcgcggcg 1980

aggcgggcgc gcccgacggc tgccccatgc gcacgggtgg gctgctcacg tgctaccgcg 2040

gcgcctcgcc cgcgccgcag cccgtcatgg cgctgcaggt ggacgtgtcc gacgctgacg 2100

cgcccttaa cctctccaag aagtcgccgt cgccgccgcg ctcctttatg ccacagatgt 2160

tggaggcgtg agatcggccc gcggagccac ttgtctccgc caccccaccc cgacctactt 2220

taaattaaat aatctagtta atagtaatcg ttcgcagatc gttttaatat attatatgtg 2280

ataatttgat taggacgcgt gagtgcgtgg ccggcgccgg acgaggagag cgcaatagac 2340

tgaaaaataa tttaagacta tataagtgta atttataaat gcaatgcagc gccggtcgcg 2400

gtgtgtttgt tgtgtgtatt gtatgtgcgt atgtgagcgt gctcagtcat ttgtttgtta 2460

atataatgac aggttgtcga actaaaataa aagaatcttg taaaaattaa aaaaaaaaaa 2520

aaaaaa 2526

<210> 20
<211> 711
<212> PRT
<213> Heliothis virescens
<400> 20

Met Arg Leu Pro Pro Leu Pro Leu Pro Asp Met Thr Val Thr Glu Cys
1 5 10 15

Gln Arg Arg Leu Leu Glu Pro Ser Ala Ala Glu Pro Pro Pro Pro Ala
20 25 30

Pro Pro Thr Asp Ser Asp Val Leu Leu Gly Arg Val Leu Ala Glu Phe
35 40 45

Asp Gly Thr Thr Val Leu Cys Arg Val Cys Gly Asp Lys Ala Ser Gly
50 55 60

Phe His Tyr Gly Val His Ser Cys Glu Gly Cys Lys Gly Phe Phe Arg
65 70 75 80

Arg Ser Ile Gln Gln Lys Ile Gln Tyr Arg Pro Cys Thr Lys Asn Gln
85 90 95

Gln Cys Ser Ile Leu Arg Ile Asn Arg Asn Arg Cys Gln Tyr Cys Arg
100 105 110

```
Leu Lys Lys Cys Ile Ala Val Gly Met Ser Arg Asp Ala Val Arg Phe
        115                 120                 125

Gly Arg Val Pro Lys Arg Glu Lys Ala Arg Ile Leu Ala Ala Met Gln
    130                 135                 140

Gln Ser Ser Thr Ser Arg Ala Asn Glu Gln Ala Ala Ala Ala Glu Leu
145                 150                 155                 160

Asp Asp Ala Pro Arg Leu Leu Ala Arg Val Val Arg Ala His Leu Asp
                165                 170                 175

Thr Cys Glu Phe Thr Arg Asp Arg Val Ala Ala Met Arg Ala Arg Ala
            180                 185                 190

Arg Asp Cys Pro Ile Tyr Ser Gln Pro Thr Leu Ala Cys Pro Leu Asn
        195                 200                 205

Pro Ala Pro Glu Leu Gln Ser Glu Lys Glu Phe Ser Gln Arg Phe Ala
    210                 215                 220

His Val Ile Arg Gly Val Ile Asp Phe Ala Gly Leu Ile Pro Gly Phe
225                 230                 235                 240

Gln Leu Leu Thr Gln Asp Asp Lys Phe Thr Leu Leu Lys Ser Gly Leu
                245                 250                 255

Phe Asp Ala Leu Phe Val Arg Leu Ile Cys Met Phe Asp Ala Pro Leu
            260                 265                 270

Asn Ser Ile Ile Cys Leu Asn Gly Gln Val Met Lys Arg Asp Ser Ile
        275                 280                 285

Gln Ser Gly Ala Asn Ala Arg Phe Leu Val Asp Ser Thr Phe Lys Phe
    290                 295                 300

Ala Glu Arg Met Asn Ser Met Asn Leu Thr Asp Ala Glu Ile Ala Leu
305                 310                 315                 320

Phe Cys Ala Ile Val Leu Ile Thr Pro Asp Arg Pro Gly Leu Arg Asn
                325                 330                 335

Val Glu Leu Val Glu Arg Met His Ala Arg Leu Lys Ala Cys Leu Gln
            340                 345                 350

Thr Val Val Ala Gln Asn Arg Pro Asp Arg Pro Gly Phe Leu Arg Glu
        355                 360                 365

Leu Met Asp Thr Leu Pro Asp Leu Arg Thr Leu Ser Thr Leu His Thr
    370                 375                 380

Glu Lys Leu Val Val Phe Arg Thr Glu His Lys Glu Leu Leu Arg Gln
385                 390                 395                 400

Gln Met Trp Thr Asp Glu Glu Gly Val Met Ser Trp Gly Asp Ser Gly
                405                 410                 415

Ala Asp Glu Ser Ala Arg Ser Pro Ile Gly Ser Val Ser Ser Ser Glu
            420                 425                 430
```

Ser Gly Glu Ala Val Gly Asp Cys Gly Thr Pro Leu Leu Ala Ala Thr
435 440 445

Leu Ala Gly Arg Arg Arg Leu Asp Ser Arg Gly Ser Val Asp Glu Glu
450 455 460

Ala Leu Gly Val Ala His Leu Ala His Asn Gly Leu Thr Val Thr Pro
465 470 475 480

Val Arg Pro Pro Pro Arg Tyr Arg Lys Leu Asp Ser Pro Thr Asp Ser
485 490 495

Gly Ile Glu Ser Gly Asn Glu Lys His Glu Arg Ile Val Gly Pro Gly
500 505 510

Ser Gly Cys Ser Ser Pro Arg Ser Ser Leu Glu Glu His Thr Asp Asp
515 520 525

Arg Arg Pro Pro Pro Ser Ala Asp Asp Met Pro Val Leu Lys Arg Val
530 535 540

Leu Glu Ala Pro Pro Leu Tyr His Thr Thr Ser Leu Met Asp Glu Ala
545 550 555 560

Tyr Lys Pro His Lys Lys Phe Arg Ala Met Arg Arg Asp Thr Gly Glu
565 570 575

Ala Glu Ala Arg Val Val Arg Pro Ala Pro Ser Thr Gln Pro Pro Gln
580 585 590

His Pro His Pro Ala Ser Pro Ala His Pro Ala His Ser Pro Arg Pro
595 600 605

Leu Arg Ala Ser Leu Ser Ser Thr His Ser Val Leu Ala Lys Ser Leu
610 615 620

Met Glu Gly Pro Arg Met Thr Pro Glu Gln Leu Lys Arg Thr Asp Ile
625 630 635 640

Ile Gln Gln Tyr Met Arg Arg Gly Glu Ala Gly Ala Pro Asp Gly Cys
645 650 655

Pro Met Arg Thr Gly Gly Leu Leu Thr Cys Tyr Arg Gly Ala Ser Pro
660 665 670

Ala Pro Gln Pro Val Met Ala Leu Gln Val Asp Val Ser Asp Ala Asp
675 680 685

Ala Pro Leu Asn Leu Ser Lys Lys Ser Pro Ser Pro Pro Arg Ser Phe
690 695 700

Met Pro Gln Met Leu Glu Ala
705 710

<210> 21
<211> 2951
<212> DNA
<213> Heliothis virescens
<400> 21
gcacgagtca aattaacgtt ttattacgag tgtgattcat aatatcttca aagacatttg 60

tgtacagcaa gggggtgtat tgtttcatct tcttataggt atatttagtt caacgacctt 120

gttactgaca ggttcatcgt tagccactga actattctat gaaagtggta tatcgtagac 180

gatagataaa ttaaatccaa tgatggagcc ctcgagagat tcagggctaa acttggaggg 240

aggttttatg tcgccgatgt caccgccgga gatgaagcca gacacggcga tgctagacgg 300

cctgcgagac gactccaccc caccccagc tttcaagaac taccccccga accatcccct 360

aagtggttct aagcacctct gttctatatg tggagataga gcgtcgggga aacattatgg 420

agtatacagt tgtgaaggtt gcaaaggttt cttcaaaagg acggtaagaa aagacttaac 480

gtacgcgtgc cgcgaagaac gcaactgcat catagacaaa cgtcagagga acagatgcca 540

gtactgcagg taccagaaat gtctcgcgtg cggcatgaag agggaagcgg tgcaggagga 600

gaggcagagg gccgccagag gtacggagga tgcacatccg agcagctcgg tgcaggtaca 660

ggagttatca atcgagcggt tgctggagat ggagtcactg gtagctgatc ccagcgaaga 720

gttccagttc cttcgtgtgg gacccgacag caatgtgcca cctaagttcc gcgccctgt 780

ctccagcctt tgtcaaatag gcaacaaaca aatagcggcg ctagtggtgt gggcgcgcga 840

catcccgcac ttcagtcagc tggagatgga agatcagatc ctgctcatca aaggctcctg 900

gaacgaactg ctgctcttcg ctatagcgtg gcggtctatg gagttcctga cagaagagcg 960

agatggcgtg gacggcactg ggaacagaac cacatcgccg ccacaactca tgtgcctcat 1020

gcctggtatg acgctgcacc gcaactcagc gctgcaggcg ggcgtggggc agatcttcga 1080

ccgcgtgctg tcggagctgt cgctgaagat gcgcaccctg cgcgtcgacc aggccgagta 1140

cgtcgcgctc aaggccatca tactgctcaa cccagatgtg aagggactga aaaacaggca 1200

agaagtggaa gttttacgag aaaagatgtt cctgtgcctg gacgagtact gccgccgctc 1260

gcgcagttcg gaggagggtc ggttcgcggc gctgctgctg cgcctgcccg cgttacgttc 1320

catttcactc aagagcttcg agcacctgtt cttcttccac ctggtggccg acaccagcat 1380

cgccggctac atccgcgacg cgctgcgcaa ccacgcgccg cccatcgaca ccaacatgat 1440

gtaaccggca cacggtgtta ctacatctgg caactaaacg aacgcgttcg acaagcgata 1500

tcgcgaccga atgtgcagct acataacgaa ctattgtcta atattgttgt ttgttgaaag 1560

actcatggca attactgagg gggatttta atcaacgatg gtataactat gctggaaata 1620

aatgaaatat gccatcggta gtaataaagt agtgctaaag cagttcaaat ctatttctca 1680

aaatggaaac tgctatgagt tttttgacaa acttttcgtc tctatttcat aaaattctat 1740

cattatacga agctaccatt tgccggtcaa aggttcgagg ttaatgtcct ggctacggtc 1800

gctaattaga gtttattata tggaaagtta gctttagcag acgtggggag ggcgtgtcga 1860

agtggttatg aacattttcg aggtcttaga ttataatctt gaaacctcac attatgtgcg 1920

tgcgtgtaac cgctgcgcga cctctaacgc gatgtaggtg tttatcgaat aaacggtacc 1980

gaatacattt gttgctataa tttgtaggaa ttttggtata cgtttcgacg gataaacctt 2040

ataatttcgg taaatttaca aatcaggcct ttatttttcg agacagcaga cttgatgtta 2100

tgttagctct aaagatcaaa tagttaacgt tggtgaaaac cttcagctat aagtacctgg 2160

gttataaata gcccttattg taaaggccct ttatgttatg ttcatgtttg agtacgagca 2220

catcaccgtt gtatgaagcg cgggcacgcg acggcagtcg tgtactaggt gtccgtgtgg 2280

gcacggtatc gacacacgac gttaggtagt gtaaatgggg ccgccttttt tattacgtag 2340

aatatgatta acgtaaatta gtcattcatt tattagggaa ttttaatcgt acaggtaggt 2400

gtccatagtg aattaatata atgttagact aatgtgtaat tttaacattt accaaatctc 2460

tcttcaaagc acgtacatta attattacgt atatttaatg gattgactat gtgtccactg 2520

agccagtgat aaaaatcatg ttgattcgta aaaacgaaaa tgaaagaaaa atatacaaaa 2580

ttatgcttca tgtatttccc gtattataca catgttggtg ctttaacatg cattatttac 2640

atattatgca tatcaaatat gcatagtaat ccttaaattt aatgttactt gataaatcct 2700

tattttcatt acgattagta tagaaagtta taaaaactgt aaatatatct ttgcatgtac 2760

aatgtatttt tattatgttt agacgagtaa cttatattgc atattaatat atacctcacc 2820

agtctcgggt ggtgggacta aacagataga atggcctaca cgctacattt ataaattata 2880

acttgtattt ctactgacag gccactttat ttttgtacct tgtgagtgcc acgcattttc 2940

caattcgcta t 2951

<210> 22
<211> 420
<212> PRT
<213> Heliothis virescens
<400> 22

Ile Asp Lys Leu Asn Pro Met Met Glu Pro Ser Arg Asp Ser Gly Leu
1 5 10 15

Asn Leu Glu Gly Gly Phe Met Ser Pro Met Ser Pro Pro Glu Met Lys
20 25 30

Pro Asp Thr Ala Met Leu Asp Gly Leu Arg Asp Asp Ser Thr Pro Pro
35 40 45

Pro Ala Phe Lys Asn Tyr Pro Pro Asn His Pro Leu Ser Gly Ser Lys
50 55 60

His Leu Cys Ser Ile Cys Gly Asp Arg Ala Ser Gly Lys His Tyr Gly
65 70 75 80

```
Val Tyr Ser Cys Glu Gly Cys Lys Gly Phe Phe Lys Arg Thr Val Arg
                85                  90                  95

Lys Asp Leu Thr Tyr Ala Cys Arg Glu Glu Arg Asn Cys Ile Ile Asp
            100                 105                 110

Lys Arg Gln Arg Asn Arg Cys Gln Tyr Cys Arg Tyr Gln Lys Cys Leu
        115                 120                 125

Ala Cys Gly Met Lys Arg Glu Ala Val Gln Glu Glu Arg Gln Arg Ala
    130                 135                 140

Ala Arg Gly Thr Glu Asp Ala His Pro Ser Ser Ser Val Gln Val Gln
145                 150                 155                 160

Glu Leu Ser Ile Glu Arg Leu Leu Glu Met Glu Ser Leu Val Ala Asp
                165                 170                 175

Pro Ser Glu Glu Phe Gln Phe Leu Arg Val Gly Pro Asp Ser Asn Val
            180                 185                 190

Pro Pro Lys Phe Arg Ala Pro Val Ser Ser Leu Cys Gln Ile Gly Asn
        195                 200                 205

Lys Gln Ile Ala Ala Leu Val Val Trp Ala Arg Asp Ile Pro His Phe
    210                 215                 220

Ser Gln Leu Glu Met Glu Asp Gln Ile Leu Leu Ile Lys Gly Ser Trp
225                 230                 235                 240

Asn Glu Leu Leu Leu Phe Ala Ile Ala Trp Arg Ser Met Glu Phe Leu
                245                 250                 255

Thr Glu Glu Arg Asp Gly Val Asp Gly Thr Gly Asn Arg Thr Thr Ser
            260                 265                 270

Pro Pro Gln Leu Met Cys Leu Met Pro Gly Met Thr Leu His Arg Asn
        275                 280                 285

Ser Ala Leu Gln Ala Gly Val Gly Gln Ile Phe Asp Arg Val Leu Ser
    290                 295                 300

Glu Leu Ser Leu Lys Met Arg Thr Leu Arg Val Asp Gln Ala Glu Tyr
305                 310                 315                 320

Val Ala Leu Lys Ala Ile Ile Leu Leu Asn Pro Asp Val Lys Gly Leu
                325                 330                 335

Lys Asn Arg Gln Glu Val Glu Val Leu Arg Glu Lys Met Phe Leu Cys
            340                 345                 350

Leu Asp Glu Tyr Cys Arg Arg Ser Arg Ser Ser Glu Glu Gly Arg Phe
        355                 360                 365

Ala Ala Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Ser Leu Lys
    370                 375                 380

Ser Phe Glu His Leu Phe Phe Phe His Leu Val Ala Asp Thr Ser Ile
385                 390                 395                 400
```

Ala Gly Tyr Ile Arg Asp Ala Leu Arg Asn His Ala Pro Pro Ile Asp
405 410 415

Thr Asn Met Met
420

<210> 23
<211> 645
<212> DNA
<213> Heliothis virescens
<400> 23
gcacgaggtg gcgaccgctt taccggcaaa cattatggag cctcgtcctg cgatggctgc 60

aagggcttct tcaggcgaag cgtcaggaaa aatcatctgt acacatgcag gttcagcaga 120

aattgtgtgg tggacaaaga caagagaaac cagtgcagat actgtaggct aaggaaatgc 180

ttcaaggcgg gcatgaagaa agaagcggtg cagaacgagc gggaccgcat caactgccga 240

agaccttcgt atgaggagcc gacgcaggcc aacgggctgt cggtagtgtc gctactcaat 300

gctgaactgc tcagtaggaa ggttattgat gagacaaaca acgtgacaga cgcggagatc 360

aacaaccgca agctggctaa gatcaacgac gtgtgcgaca gcatcaagca gcagctgctc 420

atactggtcg agtgggccaa gtacataccc gccttcactg agctgcatct tgatgatcag 480

gtggcgctgc tgcgcgcgca cgcaggcgag cacctgctgc tgggctgcgc gcgccgctcg 540

ctgcacctca gcgacatcct gctgctcggc aacaactgca tcatcaccaa gcacaatatc 600

gatggccgca tggacataga catcagcatg atcggcatgc gcgtg 645

<210> 24
<211> 215
<212> PRT
<213> Heliothis virescens
<400> 24

Ala Arg Gly Gly Asp Arg Phe Thr Gly Lys His Tyr Gly Ala Ser Ser
1 5 10 15

Cys Asp Gly Cys Lys Gly Phe Phe Arg Arg Ser Val Arg Lys Asn His
20 25 30

Leu Tyr Thr Cys Arg Phe Ser Arg Asn Cys Val Val Asp Lys Asp Lys
35 40 45

Arg Asn Gln Cys Arg Tyr Cys Arg Leu Arg Lys Cys Phe Lys Ala Gly
50 55 60

Met Lys Lys Glu Ala Val Gln Asn Glu Arg Asp Arg Ile Asn Cys Arg
65 70 75 80

Arg Pro Ser Tyr Glu Glu Pro Thr Gln Ala Asn Gly Leu Ser Val Val
85 90 95

Ser Leu Leu Asn Ala Glu Leu Leu Ser Arg Lys Val Ile Asp Glu Thr
100 105 110

Asn Asn Val Thr Asp Ala Glu Ile Asn Asn Arg Lys Leu Ala Lys Ile
115 120 125

Asn Asp Val Cys Asp Ser Ile Lys Gln Gln Leu Leu Ile Leu Val Glu
130 135 140

Trp Ala Lys Tyr Ile Pro Ala Phe Thr Glu Leu His Leu Asp Asp Gln
145 150 155 160

Val Ala Leu Leu Arg Ala His Ala Gly Glu His Leu Leu Leu Gly Cys
165 170 175

Ala Arg Arg Ser Leu His Leu Ser Asp Ile Leu Leu Leu Gly Asn Asn
180 185 190

Cys Ile Ile Thr Lys His Asn Ile Asp Gly Arg Met Asp Ile Asp Ile
195 200 205

Ser Met Ile Gly Met Arg Val
210 215

<210> 25
<211> 716
<212> DNA
<213> Heliothis virescens
<220>
<221> misc_feature
<222> (1)..(716)
<223> n is a, c, g, or t
<400> 25
gcacgagtgc aagggcttct ttagacgatt acaaagcaca gtggtgaact accaatgtcc 60

acggaataaa gcctgcgtgg tggaccgcgt caacaggaat cgctgccagt actgcagatt 120

gcagaaatgc ctcaaacttg gaatgagccg tgatgcggtg aagtttggac ggatgtcgaa 180

aaagcaacgt gagaaggtgg aagatgaggt gagattccat cgtgcacaga tgagagcgca 240

gactgacaca gcgcccgatt cagtatacga cgcccaacaa cagacgccga gctcaagcga 300

ccagttccat ggtcactaca atggctaccc gggctacgga tccccattgt cctcatacgg 360

gtacaacaac gcgggcccag cattacagtc caacatgggc ggcatacagc cccaaccgcc 420

ccaacagcag ccctacgacg tgtctgctga ctacgtggac tccaccacag cttacgagcc 480

gaaacagaac cggggattct tggaccctga ttntattagt catgcggagg gtgacatcag 540

caaagttctg gtgaagagtt tggcagaagc acatgcgaat accaacctta nactggnant 600

cattcacgag atttcagaaa gccacaaaga cgtcttccca agctttncta tactatagnt 660

cctatgacgt accgaaaggn aaaaagggaa aaaaatattt ttttttgggg gggggc 716

<210> 26
<211> 218
<212> PRT
<213> Heliothis virescens

<220>
<221> misc_feature
<222> (1)..(218)
<223> Xaa is any amino acid
<400> 26

```
His Glu Cys Lys Gly Phe Phe Arg Arg Leu Gln Ser Thr Val Val Asn
1               5                  10                  15

Tyr Gln Cys Pro Arg Asn Lys Ala Cys Val Val Asp Arg Val Asn Arg
            20                  25                  30

Asn Arg Cys Gln Tyr Cys Arg Leu Gln Lys Cys Leu Lys Leu Gly Met
        35                  40                  45

Ser Arg Asp Ala Val Lys Phe Gly Arg Met Ser Lys Lys Gln Arg Glu
    50                  55                  60

Lys Val Glu Asp Glu Val Arg Phe His Arg Ala Gln Met Arg Ala Gln
65                  70                  75                  80

Thr Asp Thr Ala Pro Asp Ser Val Tyr Asp Ala Gln Gln Gln Thr Pro
                85                  90                  95

Ser Ser Ser Asp Gln Phe His Gly His Tyr Asn Gly Tyr Pro Gly Tyr
            100                 105                 110

Gly Ser Pro Leu Ser Ser Tyr Gly Tyr Asn Asn Ala Gly Pro Ala Leu
        115                 120                 125

Gln Ser Asn Met Gly Gly Ile Gln Pro Gln Pro Pro Gln Gln Gln Pro
    130                 135                 140

Tyr Asp Val Ser Ala Asp Tyr Val Asp Ser Thr Thr Ala Tyr Glu Pro
145                 150                 155                 160

Lys Gln Asn Arg Gly Phe Leu Asp Pro Asp Xaa Ile Ser His Ala Glu
                165                 170                 175

Gly Asp Ile Ser Lys Val Leu Val Lys Ser Leu Ala Glu Ala His Ala
            180                 185                 190

Asn Thr Asn Leu Xaa Leu Xaa Xaa Ile His Glu Ile Ser Glu Ser His
        195                 200                 205

Lys Asp Val Phe Pro Ser Phe Xaa Ile Leu
    210                 215
```

<210> 27
<211> 5206
<212> DNA
<213> Drosophila melanogaster
<400> 27

```
ggctcgccca ttggagggcc cctgtcctgt ggcagcagct tgcccagctt ccaggagacc     60

tactccttga agtacaacag cagcagcggt agcagccccc agcaggcgtc ctcctcctcc    120

accgccgccc ccacgcccac tgaccaggtg ctgaccctca agatggacga ggactgcttc    180

ccgcctctgt ccggcggctg gagtgccagt ccgcccgccc cctcccagct ccagcagctg    240
```

cacaccctgc agtctcaggc ccagatgtcg catcccaaca gcagcaacaa cagcagcaac 300

aacgcgggca acagccacaa caacagtggg ggctacaact accacggcca cttcaatgcc 360

atcaatgcca gcgccaatct gtcgcccagc tcctcggcca gttccctcta cgaatataat 420

ggtgtttccg cagcggacaa cttctacgga caacagcagc agcagcaaca gcaaagctat 480

cagcaacata actacaactc gcacaatggc gagcgttact cgctgcccac gtttcccacg 540

atttcggagc tggctgcggc cactgctgct gtcgaagctg cggcggcggc cacagtgggc 600

ggtccgccgc cagtacgccg agcatcgctg ccggttcagc gaaccgttct gccagccggc 660

tccacggcgc agagccccaa gctggccaag atcacactga accagcggca ctcacatgcc 720

catgcccatg ccctacagct caactcggca cccaattcgg cggcaagttc gccagcgagt 780

gcggatctgc aggcgggccg tttgctccag gctccgtcgc agctgtgtgc tgtttgtggc 840

gacaccgccg cctgtcagca ttatggagtg cgaacctgcg agggatgcaa gggatttttc 900

aagcggaccg tgcagaaggg ctccaagtac gtctgcctgg cggacaagaa ttgcccggtg 960

gacaagaggc gccgcaaccg ttgccagttc tgccggttcc agaagtgcct ggtcgtaggc 1020

atggtcaagg aagtggtgcg caccgactcg ttgaagggtc gccgcggaag actgccctca 1080

aaaccgaaat cgccccagga gtcgccacca tcaccaccca tctcgttgat cacggccctg 1140

gtgcgcagcc atgtcgacac gactccggat ccctcgtgcc tggactatag ccactacgag 1200

gagcagtcga tgagcgaggc agataaggtg caacagtttt accagctgct gaccagctcc 1260

gtggacgtga tcaaacagtt cgccgagaag attcccggct acttcgatct cctgccggag 1320

gatcaggagc tgctcttcca gagcgcatcg ctggaactgt tcgtcctgcg gctggcctat 1380

cgcgccagga tcgatgacac caagctgatc ttctgcaacg gcacggtgct ccaccgcacc 1440

cagtgcctgc gctctttcgg cgagtggctc aacgacatca tggagttcag ccgcagcctg 1500

cacaacctgg agatcgacat ctccgccttc gcctgcctct gtgccctaac cctcatcaca 1560

gaacgccatg gcctgcggga gccgaagaag gtggagcagc tccagatgaa gatcattggc 1620

agtctgcgcg accacgtcac ctacaatgcc gaggcccaga agaagcagca ctacttcagc 1680

cgcctgctgg gcaagctgcc ggagctgagg tccctgagtg tccagggact gcagaggatc 1740

ttctacctga agctggagga cctggtaccc gcaccagctc tcatcgagaa catgttcgtc 1800

accacattgc ccttctagag gcgatcatca agcgtatcat cacaacttgc ttccttaaaa 1860

ctagccccta agtatgcctc ctaggatata cagagaaagg accccatagg acggacgcaa 1920

ctagctttag tagaaccctg aaataaataa atctcacaac agcaaaaaca aaaccgaacc 1980

gaacagaaat gaagcgaata gcagacccag gccatatctt tagtgtagag ctaggtagtt 2040

```
accgggacag cccggctcct tcgataatta cggacatgca tatttgagag ggggtttcca   2100
gtgcacagcc tatggctcct gcgtgactcg tcagcacgcg agctccaact tgttgacgtt   2160
aattgttaaa ttgtttaatt tcaactgtca aaaccggaat caacggccgg gcacgcaatg   2220
gcaacacttt ctatccccgg acttcgaagc ctgctcaaca ttcggcacta cggacggaca   2280
aacaacggac agaaacagaa ctcactcttg ctctcttgcc ttttactaac ttctagtcaa   2340
ttgatttagg cgaatcaaat aaataaataa ataaaataag ggcgtgcagc agtagtgtta   2400
tataatttat atccagaccc cagcggttct cttcaaggaa atcccccaat gagttgcaca   2460
aattgggata aagtacgata gcctattatt cttatatttc ttttaaaagc tcgaagatag   2520
atgagaactg tgttggaaat tccactatca tatcatatag ttgctataag ccgtgcttgc   2580
cctaagctaa gttagacccg cataaagttg atagcccaac caagtatttc ggttatttcc   2640
tagactaagg tcctaatagt tataggctaa gactattctg ttcgattatc aatgcaccaa   2700
acagtgcaca atgagagtat aagtaccttc ttgtgatgat tgtgtctgac acagagagag   2760
ttgcacacaa gcacacaaac tagccgataa gttactaaat acgatctaat atctaatata   2820
tataatataa tataatatat ataagtccaa gtattcggaa atccaagaac ccttgcataa   2880
ccgcagttcg tacgttccaa acgagaaaag aactttattt aatcctagac ccactcatct   2940
aagcttctaa agaatcgtat gtggatcgtt ggatctgtct ctctatatat gtgtgtgtgt   3000
gttatctcga tagaaaaccc ctctatgtga ttttgtgata gattggcatt gaactctata   3060
tatttatata tatatatgtc tataatatat atacacgcat aaatatatat ttttatgtct   3120
aacttttgta tggtttattt tatacgtacc acttttcttt gataacaaaa agtaaaaaac   3180
tcgttagata gcaaatattt caaaggtatg ttacgaggac ttttcaaagg tatgttcgag   3240
gacttttcaa agtaccagtc tttagcgact ttccaattaa cgttcgtatt aacgaaagac   3300
agattttcta tgtgttaaat tgaaagactt ctataactat aactaaatgc aagctaagag   3360
caaaaacaca aatccacaaa tccccaaagt gaataacata tctcttcaag ctttcgagtg   3420
ctcggaacac gtagaaccga aacccaagtg ttactaaatc catttaataa tcgcaagccg   3480
ggggcgtcgg cgtggttaat acgttctcat tacctataca atttagatag atcattatta   3540
aattattgta catgtagcac atgaaatgtt cgacaactag attttgtacc atcttaaaga   3600
agaacctagg ccaagctaaa ctaagtataa actatgatct gcatgcggct agctgtagct   3660
atgagaaata tacctgcgtg gatctaagtg aaatgggaca ctttgaattt agatatgaaa   3720
cgttctaaac gcgacgacta actctcccaa ctgcgaactc taccaattaa gagaaattcc   3780
cagaaaatgt gtcaggattt caaagcgtcc catctcactt gaacccaccc aatcaacaaa   3840
tacaaatcct agggaagttg agaggttcag caaccataga gcaatatttc ataagaaaac   3900
```

gcaccttaaa ttaccgaaaa acatagatta acctgatctt gtaacgtttg ggagcgataa 3960

taagccagga ttaaacagga acagttaggt gaccaaatca gttcgaaacg aagatgatag 4020

atagttcggt tcgaaaccct aaacgcgatg ccattttagc cgttacaaca ttggatatca 4080

atcatgcaca tgaatatgaa tatgaatatg aatattatag agatatatct agctatagga 4140

acctactttg tacctacacg acatggaaac atcaaaccta catgcatatt tacacacata 4200

tattttgaat agagcgacga cttttacaag ttgcgtagaa agctatagct atagcttgat 4260

atggccatcc cagagcgagc atatacatat attttgggtt attgttcttt tgtaatttta 4320

taaatgcata catatttatg tactacgtga atgtcaagtg tggattcata tttttgagat 4380

acagctacaa aaacgaaaca aaagaaaata aaacaaaaca gaagagtaaa cggtgaaatt 4440

ttttcgatga aacaatttta aatgagaact ttttaatatt gctattaaag gatatacata 4500

tacacactaa catacaatat atattttact atgtaacgga tagaattaag ctagatgcag 4560

cgcataaagc tttatacaac aaattgaaaa gcaacagaag aaattggcac aaattaaatt 4620

tatatagcat aattagacgt ccttcgcaag ataatgttat tgctaataag agcgtcaatc 4680

ggtacatcgg gcgctatttc ccactacacc cccaaccaca caatagataa cctaagcaat 4740

gtatgtacat tagctatgta tatccagccc acttatgcgc ctactactag aaatgcagaa 4800

agcagaaaga gaggtgaaac ctatagacgc tatcacaaat gtctatctga tagacatcgg 4860

tactaccaat gctatattgc cagttgtgta atttactctt atttgatcgt ttcatttacc 4920

agttaagaac ccaaatcata taagtgttat gatggaagaa ctataacttg caattcaatt 4980

aactctgcat acgataacaa gcaaagcgaa tcatttcatt tcgatttaat ctttaattat 5040

atatacttaa acgatgtaag cccaaaacaa acgtttttc tatatctgtc ttttgagcaa 5100

attagttata cgcaaaacca aaccgtattt acataaatgt atacaaaaca aatcgtatat 5160

tttcattggt ttgaaataaa tacataaaac aaaaaaaaaa aaaaaa 5206

<210> 28
<211> 551
<212> PRT
<213> Drosophila melanogaster
<400> 28

Met Asp Glu Asp Cys Phe Pro Pro Leu Ser Gly Gly Trp Ser Ala Ser
1 5 10 15

Pro Pro Ala Pro Ser Gln Leu Gln Gln Leu His Thr Leu Gln Ser Gln
20 25 30

Ala Gln Met Ser His Pro Asn Ser Ser Asn Asn Ser Ser Asn Asn Ala
35 40 45

```
Gly Asn Ser His Asn Asn Ser Gly Gly Tyr Asn Tyr His Gly His Phe
    50                  55                  60

Asn Ala Ile Asn Ala Ser Ala Asn Leu Ser Pro Ser Ser Ser Ala Ser
65                  70                  75                  80

Ser Leu Tyr Glu Tyr Asn Gly Val Ser Ala Ala Asp Asn Phe Tyr Gly
                85                  90                  95

Gln Gln Gln Gln Gln Gln Gln Gln Ser Tyr Gln Gln His Asn Tyr Asn
            100                 105                 110

Ser His Asn Gly Glu Arg Tyr Ser Leu Pro Thr Phe Pro Thr Ile Ser
        115                 120                 125

Glu Leu Ala Ala Ala Thr Ala Ala Val Glu Ala Ala Ala Ala Ala Thr
    130                 135                 140

Val Gly Gly Pro Pro Pro Val Arg Arg Ala Ser Leu Pro Val Gln Arg
145                 150                 155                 160

Thr Val Leu Pro Ala Gly Ser Thr Ala Gln Ser Pro Lys Leu Ala Lys
                165                 170                 175

Ile Thr Leu Asn Gln Arg His Ser His Ala His Ala His Ala Leu Gln
            180                 185                 190

Leu Asn Ser Ala Pro Asn Ser Ala Ala Ser Ser Pro Ala Ser Ala Asp
        195                 200                 205

Leu Gln Ala Gly Arg Leu Leu Gln Ala Pro Ser Gln Leu Cys Ala Val
    210                 215                 220

Cys Gly Asp Thr Ala Ala Cys Gln His Tyr Gly Val Arg Thr Cys Glu
225                 230                 235                 240

Gly Cys Lys Gly Phe Phe Lys Arg Thr Val Gln Lys Gly Ser Lys Tyr
                245                 250                 255

Val Cys Leu Ala Asp Lys Asn Cys Pro Val Asp Lys Arg Arg Arg Asn
            260                 265                 270

Arg Cys Gln Phe Cys Arg Phe Gln Lys Cys Leu Val Val Gly Met Val
        275                 280                 285

Lys Glu Val Val Arg Thr Asp Ser Leu Lys Gly Arg Arg Gly Arg Leu
    290                 295                 300

Pro Ser Lys Pro Lys Ser Pro Gln Glu Ser Pro Pro Ser Pro Pro Ile
305                 310                 315                 320

Ser Leu Ile Thr Ala Leu Val Arg Ser His Val Asp Thr Thr Pro Asp
                325                 330                 335

Pro Ser Cys Leu Asp Tyr Ser His Tyr Glu Glu Gln Ser Met Ser Glu
            340                 345                 350

Ala Asp Lys Val Gln Gln Phe Tyr Gln Leu Leu Thr Ser Ser Val Asp
        355                 360                 365
```

Val Ile Lys Gln Phe Ala Glu Lys Ile Pro Gly Tyr Phe Asp Leu Leu
370 375 380

Pro Glu Asp Gln Glu Leu Leu Phe Gln Ser Ala Ser Leu Glu Leu Phe
385 390 395 400

Val Leu Arg Leu Ala Tyr Arg Ala Arg Ile Asp Asp Thr Lys Leu Ile
405 410 415

Phe Cys Asn Gly Thr Val Leu His Arg Thr Gln Cys Leu Arg Ser Phe
420 425 430

Gly Glu Trp Leu Asn Asp Ile Met Glu Phe Ser Arg Ser Leu His Asn
435 440 445

Leu Glu Ile Asp Ile Ser Ala Phe Ala Cys Leu Cys Ala Leu Thr Leu
450 455 460

Ile Thr Glu Arg His Gly Leu Arg Glu Pro Lys Lys Val Glu Gln Leu
465 470 475 480

Gln Met Lys Ile Ile Gly Ser Leu Arg Asp His Val Thr Tyr Asn Ala
485 490 495

Glu Ala Gln Lys Lys Gln His Tyr Phe Ser Arg Leu Leu Gly Lys Leu
500 505 510

Pro Glu Leu Arg Ser Leu Ser Val Gln Gly Leu Gln Arg Ile Phe Tyr
515 520 525

Leu Lys Leu Glu Asp Leu Val Pro Ala Pro Ala Leu Ile Glu Asn Met
530 535 540

Phe Val Thr Thr Leu Pro Phe
545 550

<210> 29
<211> 4778
<212> DNA
<213> Drosophila melanogaster
<400> 29
gaaaaaacaa acattttgct acttcgtcgc aggcgggact gtgttgcgtc gtgtgatcgc 60

tagagcggtt gtggaatcgg attcgagcgc aaaacaccgt tcatgctgtg aaaaatccga 120

tatttgtcgt gcaataattt cctcgattgg catcaagtgg cttccagtcg ggtacatatt 180

gcacaagaaa tgttatacgc ataatgtgca cgcaaattaa acgaattctc tatgaaaatg 240

tgactagaat gtgagtcgaa caaaacgagt aaaacgtgaa atcccaactg gcttttgggt 300

aacaaatctt atcaacacag caacggaaat acattaaaat cttgatagac tgagaaaggg 360

acaattggaa tacttttagt tatttttaaa tgttttacaa cacaatggaa ctgcatcaac 420

gacacctctc aaacttttac aaattgcaca actgagaaat agtctttgat aaataaataa 480

aatataagaa atcgctactg aaacaagatg ccaaacatgt ccagcatcaa agcggagcag 540

caaagcggtc ctcttggagg aagtagcggc tatcaagtac cggtcaacat gtgcaccacc 600

```
acagtcgcga atacgacgac cactttggga agctccgccg ggggagccac tggctcccgg    660
cacaacgtct ccgtgacaaa catcaagtgc gaactagacg aactaccgtc accgaacggc    720
aacatggtgc cggttatcgc aaactacgtt cacggtagct tgcgcattcc actcagtgga    780
cattcaaatc atagggagtc cgattcggag gaggagctgg caagtattga gaacttgaag    840
gttcggcgaa ggacggcggc ggacaaaaat ggtcctcgtc caatgtcctg ggagggcgag    900
ctgagcgata ctgaggtcaa cgggggcgaa gagctgatgg aaatggagcc aacaattaag    960
agtgaggtgg tccctgctgt tgcaccccca caacccgtct gcgcactaca accgataaaa   1020
acagagctag agaacattgc aggcgagatg cagattcaag ggaagtgtta cccccagtcc   1080
aacacacaac atcacgctgc cacaaaatta aaagtggccc cgacgcaaag tgatccgatc   1140
aatctcaagt tcgaaccgcc tctgggagac aattctccgc tactggctgc acgtagcaag   1200
tccagcagtg gaggccacct accactgcca acgaatccca gtcccgactc cgccatacat   1260
tccgtctaca cgcacagctc cccctcgcag tcgcctctga cgtcgcgcca cgccccctac   1320
actccgtctc tgagccgcaa caacagcgac gcctcgcaca gtagctgcta cagctatagc   1380
tccgaattca gtcccacaca ctcgcccatt caagcgcgtc atgccccacc cgccggcacg   1440
ctctatggca accaccatgg tatttaccgc cagatgaagg tggaagcctc atccactgtg   1500
ccgtccagtg ggcaggaggc gcagaacctg agtatggact ctgcctctag caatctggat   1560
acagtgggct taggatcttc gcaccccgca tctccggcgg gcatatcacg tcagcagttg   1620
atcaactcgc cctgccccat ctgcggtgac aagatcagcg gatttcatta ctttattttc   1680
tcctgcgagt cttgcaaggg cttcttcaag cgcaccgtgc aaaatcgcaa gaactacgtg   1740
tgcgtgcgtg gtggaccatg tcaggtcagc atttccacgc gcaagaaatg tccagcctgc   1800
cgcttcgaga agtgtctgca gaagggaatg aaactagaag cgattcggga ggaccgaacc   1860
cgtggcggcc gctccacata ccagtgctcc tacacgctgc ccaactcaat gcttagtccg   1920
ctgcttagtc ctgatcaagc ggcagcagct gccgccgcag cagcagtggc aagtcagcag   1980
cagccgcacc agcgactaca tcaactaaat ggatttggag gtgtacccat tccctgctct   2040
acttctcttc cagccagccc tagtttggca ggaacttcgg tcaagtcgga agagatggcg   2100
gagacgggca agcaaagcct ccgaacggga agcgtaccac cactactgca ggaaatcatg   2160
gatgtagagc atctgtggca gtacaccgat gcagagctgg cccgcatcaa ccaaccactg   2220
tccgcattcg cctctggcag ctcttcgtcg tcgtcatcgt caggtacatc ctcaggcgcc   2280
catgcacaac tcaccaatcc actactggct agtgctggtc tctcgtccaa tggcgagaat   2340
gccaatcctg atcttatcgc tcatctctgc aacgtggctg atcaccgtct ttataaaatc   2400
```

gtcaaatggt gcaagagctt gccgcttttt aagaacattt cgatcgatga ccaaatctgc 2460

ttgctcatta actcgtggtg cgagctgttg ctcttctcct gctgttttag atcaattgat 2520

actcctggag agattaaaat gtcacaaggc aggaagataa ccctatcgca ggccaaatca 2580

aatggcttgc agacttgcat tgaacggatg ctcaacctaa cagatcacct gaggcgattg 2640

cgcgttgatc gctacgaata tgttgccatg aaagttattg tgctgttgca gtcagatacg 2700

acagagttac aggaagcggt aaaggtgcgc gagtgtcagg aaaaagcttt gcagagcttg 2760

caagcttaca ccctggcgca ttatcctgac acgcaatcca agtttgggga gcttttgcta 2820

cgcattcctg atttgcagcg aacgtgccag cttggcaagg agatgttgac gatcaagact 2880

cgcgatggag ctgatttcaa tttgctaatg gagcttttgc gcggagagca ttgacaattg 2940

ataactaaga cggaaatctt ttaccattgg caaaacaagt ttcacatatt tagtattaga 3000

tatatatatt ctatagataa gatccttact gtaagttctg aaaacatgtg cctaaaaacc 3060

aaagccacga tagcagtcac atcaggccca ctggtcgaga ttaaatccaa gagcaagatt 3120

gccaaatttt tacaccaata tatattttga tatgagccat gtgcagggcc tcagatcgct 3180

gttgttgtcg gctaaagttt cagtaaggaa agtatatatt gattttgcta tttatacata 3240

tttgacttat gtatagtgta aactaaagca cacatggaaa atgaaaagac taaacaaatt 3300

tatttaaaaa ttacttttac tattatagaa aaaggggaaa aataaaaaac acaaaggcag 3360

agaagaaaat ttagttacaa caggtagcga catttttata ttttcttata taaggaaata 3420

ttcaatgtat tttaaatata aagccaaacc cgatttggtt tggggaagag ctactgaaat 3480

ttttgatatc tatatattca ttactagaag acgaatgaat gtatccaatg tttaaatgtt 3540

gtagcgttta gttttagtgc aatttcacac atgtctacat acatgaatat tcagcgagat 3600

atgtctgcaa actattataa agcaaaagac cactcgaaat cgccatcact ggttggctaa 3660

gactattcca gttatgctgt ttgttgcata aaaaaccaca actacgtaca tcaataaaat 3720

gtataatttt ttattggagt tttagatttg tattaacttc ttccttataa ttacgattat 3780

tattattatt actaatttta tgattattgt gtaacactga cttaaatagc tgaaacaaat 3840

cctgcaacag gatttaaaac acctgaatac acaaaacatt ataacatgaa tacattttgc 3900

ttatggccta gatagtttga tatgtacttt gatatgtatg catgtgtcta tatgtgagtg 3960

cgtccataca aattcctgtc ccaccagaaa aatcacacgc aataaaaaat tccaaaatac 4020

taagctcgta tctacaaaga aagattaaaa gacaaattga tgaataggaa tatgttgccg 4080

gaagtccaag agatttggct gaaagtatcg acaaattttc aacacatcgt tcatggatat 4140

tgtgctaaca ctctcagttt gaaaatcatt ttctgttaaa ctttctatat aataagttct 4200

ccattcgatt ttgtatttac aatttgtttc tttaattttc ctttatcagt tgtatctatg 4260

aaacatgagg atcacagttc atattgatcg tgttcttctg ccgtacaccg cttctgtccg 4320

ttaatgtaaa ccataagtat aaatgaaatt agttaaatgt ttatttataa ataaagcgct 4380

ataataaatt tcaatacatt tatcatagtt aactgattaa gaccactgaa atcaaaaata 4440

ttttatttac taagcaaagc acacgcaaac aatttataat gtttattacg ttaacaacaa 4500

actcattttt aataattctt tatgaataca caaagttacg caattttccc tctaggcgca 4560

ttgcttaaat agttaaagaa aaataataaa cccatagcgc aatatttaat gtaaaacagt 4620

tttccttgcg tgtgatgttt cctctagcta cgtacaaatt catcatttat taaatttaaa 4680

actcaatttt gcttttaaat aaatttaata agtaaaattc aacaataatt gatatacaat 4740

tgtcaatgca atattttgta ataaaaatgc gaaaaatc 4778

<210> 30
<211> 808
<212> PRT
<213> Drosophila melanogaster
<400> 30

Met Pro Asn Met Ser Ser Ile Lys Ala Glu Gln Gln Ser Gly Pro Leu
1 5 10 15

Gly Gly Ser Ser Gly Tyr Gln Val Pro Val Asn Met Cys Thr Thr Thr
20 25 30

Val Ala Asn Thr Thr Thr Thr Leu Gly Ser Ser Ala Gly Gly Ala Thr
35 40 45

Gly Ser Arg His Asn Val Ser Val Thr Asn Ile Lys Cys Glu Leu Asp
50 55 60

Glu Leu Pro Ser Pro Asn Gly Asn Met Val Pro Val Ile Ala Asn Tyr
65 70 75 80

Val His Gly Ser Leu Arg Ile Pro Leu Ser Gly His Ser Asn His Arg
85 90 95

Glu Ser Asp Ser Glu Glu Glu Leu Ala Ser Ile Glu Asn Leu Lys Val
100 105 110

Arg Arg Arg Thr Ala Ala Asp Lys Asn Gly Pro Arg Pro Met Ser Trp
115 120 125

Glu Gly Glu Leu Ser Asp Thr Glu Val Asn Gly Gly Glu Glu Leu Met
130 135 140

Glu Met Glu Pro Thr Ile Lys Ser Glu Val Val Pro Ala Val Ala Pro
145 150 155 160

Pro Gln Pro Val Cys Ala Leu Gln Pro Ile Lys Thr Glu Leu Glu Asn
165 170 175

Ile Ala Gly Glu Met Gln Ile Gln Gly Lys Cys Tyr Pro Gln Ser Asn
180 185 190

Thr Gln His His Ala Ala Thr Lys Leu Lys Val Ala Pro Thr Gln Ser
195 200 205

Asp Pro Ile Asn Leu Lys Phe Glu Pro Pro Leu Gly Asp Asn Ser Pro
210 215 220

Leu Leu Ala Ala Arg Ser Lys Ser Ser Ser Gly Gly His Leu Pro Leu
225 230 235 240

Pro Thr Asn Pro Ser Pro Asp Ser Ala Ile His Ser Val Tyr Thr His
245 250 255

Ser Ser Pro Ser Gln Ser Pro Leu Thr Ser Arg His Ala Pro Tyr Thr
260 265 270

Pro Ser Leu Ser Arg Asn Asn Ser Asp Ala Ser His Ser Ser Cys Tyr
275 280 285

Ser Tyr Ser Ser Glu Phe Ser Pro Thr His Ser Pro Ile Gln Ala Arg
290 295 300

His Ala Pro Pro Ala Gly Thr Leu Tyr Gly Asn His His Gly Ile Tyr
305 310 315 320

Arg Gln Met Lys Val Glu Ala Ser Ser Thr Val Pro Ser Ser Gly Gln
325 330 335

Glu Ala Gln Asn Leu Ser Met Asp Ser Ala Ser Ser Asn Leu Asp Thr
340 345 350

Val Gly Leu Gly Ser Ser His Pro Ala Ser Pro Ala Gly Ile Ser Arg
355 360 365

Gln Gln Leu Ile Asn Ser Pro Cys Pro Ile Cys Gly Asp Lys Ile Ser
370 375 380

Gly Phe His Tyr Phe Ile Phe Ser Cys Glu Ser Cys Lys Gly Phe Phe
385 390 395 400

Lys Arg Thr Val Gln Asn Arg Lys Asn Tyr Val Cys Val Arg Gly Gly
405 410 415

Pro Cys Gln Val Ser Ile Ser Thr Arg Lys Lys Cys Pro Ala Cys Arg
420 425 430

Phe Glu Lys Cys Leu Gln Lys Gly Met Lys Leu Glu Ala Ile Arg Glu
435 440 445

Asp Arg Thr Arg Gly Gly Arg Ser Thr Tyr Gln Cys Ser Tyr Thr Leu
450 455 460

Pro Asn Ser Met Leu Ser Pro Leu Leu Ser Pro Asp Gln Ala Ala Ala
465 470 475 480

Ala Ala Ala Ala Ala Ala Val Ala Ser Gln Gln Gln Pro His Gln Arg
485 490 495

Leu His Gln Leu Asn Gly Phe Gly Gly Val Pro Ile Pro Cys Ser Thr
500 505 510

Ser Leu Pro Ala Ser Pro Ser Leu Ala Gly Thr Ser Val Lys Ser Glu
515 520 525

Glu Met Ala Glu Thr Gly Lys Gln Ser Leu Arg Thr Gly Ser Val Pro
530 535 540

Pro Leu Leu Gln Glu Ile Met Asp Val Glu His Leu Trp Gln Tyr Thr
545 550 555 560

Asp Ala Glu Leu Ala Arg Ile Asn Gln Pro Leu Ser Ala Phe Ala Ser
565 570 575

Gly Ser Ser Ser Ser Ser Ser Ser Ser Gly Thr Ser Ser Gly Ala His
580 585 590

Ala Gln Leu Thr Asn Pro Leu Leu Ala Ser Ala Gly Leu Ser Ser Asn
595 600 605

Gly Glu Asn Ala Asn Pro Asp Leu Ile Ala His Leu Cys Asn Val Ala
610 615 620

Asp His Arg Leu Tyr Lys Ile Val Lys Trp Cys Lys Ser Leu Pro Leu
625 630 635 640

Phe Lys Asn Ile Ser Ile Asp Asp Gln Ile Cys Leu Leu Ile Asn Ser
645 650 655

Trp Cys Glu Leu Leu Leu Phe Ser Cys Cys Phe Arg Ser Ile Asp Thr
660 665 670

Pro Gly Glu Ile Lys Met Ser Gln Gly Arg Lys Ile Thr Leu Ser Gln
675 680 685

Ala Lys Ser Asn Gly Leu Gln Thr Cys Ile Glu Arg Met Leu Asn Leu
690 695 700

Thr Asp His Leu Arg Arg Leu Arg Val Asp Arg Tyr Glu Tyr Val Ala
705 710 715 720

Met Lys Val Ile Val Leu Leu Gln Ser Asp Thr Thr Glu Leu Gln Glu
725 730 735

Ala Val Lys Val Arg Glu Cys Gln Glu Lys Ala Leu Gln Ser Leu Gln
740 745 750

Ala Tyr Thr Leu Ala His Tyr Pro Asp Thr Gln Ser Lys Phe Gly Glu
755 760 765

Leu Leu Leu Arg Ile Pro Asp Leu Gln Arg Thr Cys Gln Leu Gly Lys
770 775 780

Glu Met Leu Thr Ile Lys Thr Arg Asp Gly Ala Asp Phe Asn Leu Leu
785 790 795 800

Met Glu Leu Leu Arg Gly Glu His
805

<210> 31
<211> 2561
<212> DNA

<213> Drosophila melanogaster
<400> 31
ctacgcaaaa taaaacgtac atgaaatgtt attagaaatg gatcagcaac aggcgaccgt 60

acagtttata tcgtcgctga atatatcgcc gttcagcatg cagctggagc agcagcagca 120

gccctccagt cccgctctgg ccgccggtgg caacagcagc aacaacgcgg ccagcggtag 180

caacaacaac agcgccagcg gcaacaacac cagcagcagc agcaacaaca acaacaataa 240

caacgacaat gatgcacacg ttctaacgaa attcgagcac gaatacaatg cctacacgtt 300

gcagttggcc ggaggcggtg ggagtggcag cggcaatcag cagcaccaca gcaaccacag 360

caaccacggc aaccaccacc agcagcagca gcaacaacag caacagcagc agcaacatca 420

gcagcagcag caagaacact accagcagca acagcagcag aatatcgcca acaatgccaa 480

tcaattcaac tcctcgtcct actcgtatat atacaatttc gattcacagt atatattccc 540

tacaggctac caggacacca cctcctcaca ctcgcaacag agcggaggag gcggtggcgg 600

cggcggtggc aacctgctaa acggcagctc cggcggcagc tccgccggcg gtggctacat 660

gctgctcccc caggcggcca gctccagtgg caataatggc aacccgaatg ccggccacat 720

gtcctccggt tccgtgggca atggcagcgg aggcgctggc aatggcggag cgggtggcaa 780

ctccggtccc ggcaatccca tgggcggcac gagcgccacg ccgggacacg gcggcgaggt 840

gatcgacttc aagcacctgt tcgaggagct ttgccccgtg tgtggcgaca aggtgagcgg 900

ctaccactac ggcctgctca cctgcgagtc ctgcaagggc ttcttcaagc gcaccgtgca 960

gaacaagaag gtctacacct gtgtggcgga gcggtcgtgc cacatcgaca agacgcagcg 1020

caagcggtgt ccctactgcc gattccagaa gtgcctcgag gtgggcatga agctagaggc 1080

tgttcgagcg gatagaatgc gtggtggacg caacaaattc ggacccatgt acaaacggga 1140

tcgcgcgcgg aagttgcaag tgatgcggca gcggcagttg gcgctgcaag cgctgcgcaa 1200

ctcgatgggt ccggacatca agccaacgcc gatctcgccg ggctaccagc aagcatatcc 1260

aaatatgaac attaagcagg aaattcaaat acctcaggta tcctcactca cccaatctcc 1320

ggactcgtcg cccagcccca tagcaattgc gttgggacag gtgaacgcga gcacgggcgg 1380

tgttatagcc acgcccatga acgccggcac tggcggcagt gggggcggtg gtctgaacgg 1440

accaagttcc gtgggcaacg gcaatagcag caacggcagc agcaacggca acaacaacag 1500

cagcacgggc aacggaacgt ccggaggagg aggtggcaat aatgcgggcg gcggaggagg 1560

aggaaccaat tccaacgatg gcctgcatcg caacggcggc aatgacagca gcagttgcca 1620

cgaggctgga ataggatctc tgcagaacac ggccgactcg aaattgtgct tcgattctgg 1680

cacacatcca tcgagcacag ccgacgcgct aatcgagcca ttaagagtct caccgatgat 1740

tcgtgaattt gtgcaatcga ttgacgatcg ggaatggcag acgcaactgt ttgccctgct 1800

gcagaagcaa acctacaacc aggtggaagt ggatctcttc gagctgctga tgtgcaaagt 1860

gctcgaccag aatttgttct cgcaagtaga ctgggcacgg aacaccgtct tcttcaagga 1920

tctgaaggtc gacgaccaaa tgaagctgct gcagcattcc tggtcggaca tgcttgttct 1980

ggatcacctg catcatcgaa tccataacgg cctgcccgac gagacgcaac tgaacaatgg 2040

tcaggtgttc aatctgatga gtctgggttt gttgggagtg ccacaacccg gcgattactt 2100

caacgagctg cagaacaagc tgcaggacct gaaattcgat atgggcgact atgtctgcat 2160

gaaattccta atcctgttga atccaagtgt acggggtatt gtcaaccgga agaccgtatc 2220

cgagggacat gataatgtgc aagccgcttt gctggactac accctcacct gctatccgtc 2280

agtgaatgac aaattcagag ggctagttaa catcttaccg gaaatccatg ccatggccgt 2340

tcgcggcgag gatcacctga tcacctgtac accaagcact gtgccggcag tgcgcccacc 2400

caaacgctgc tcatggagat gctgcacgcc aagcgcaagg gatagaggtc gggagaacgt 2460

gacacggaat acttaatcat ttatgaaatg taaataacaa ggcgggaagg ccctcggggc 2520

aaccgggaca tggaaggcga acgaaggata cagcagaatt c 2561

<210> 32
<211> 816
<212> PRT
<213> Drosophila melanogaster
<400> 32

Met Leu Leu Glu Met Asp Gln Gln Gln Ala Thr Val Gln Phe Ile Ser
1 5 10 15

Ser Leu Asn Ile Ser Pro Phe Ser Met Gln Leu Glu Gln Gln Gln Gln
20 25 30

Pro Ser Ser Pro Ala Leu Ala Ala Gly Gly Asn Ser Ser Asn Asn Ala
35 40 45

Ala Ser Gly Ser Asn Asn Asn Ser Ala Ser Gly Asn Asn Thr Ser Ser
50 55 60

Ser Ser Asn Asn Asn Asn Asn Asn Asn Asp Asn Asp Ala His Val Leu
65 70 75 80

Thr Lys Phe Glu His Glu Tyr Asn Ala Tyr Thr Leu Gln Leu Ala Gly
85 90 95

Gly Gly Gly Ser Gly Ser Gly Asn Gln Gln His His Ser Asn His Ser
100 105 110

Asn His Gly Asn His His Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
115 120 125

Gln Gln His Gln Gln Gln Gln Gln Glu His Tyr Gln Gln Gln Gln Gln
130 135 140

Gln Asn Ile Ala Asn Asn Ala Asn Gln Phe Asn Ser Ser Ser Tyr Ser
145 150 155 160

Tyr Ile Tyr Asn Phe Asp Ser Gln Tyr Ile Phe Pro Thr Gly Tyr Gln
165 170 175

Asp Thr Thr Ser Ser His Ser Gln Gln Ser Gly Gly Gly Gly Gly Gly
180 185 190

Gly Gly Gly Asn Leu Leu Asn Gly Ser Ser Gly Gly Ser Ser Ala Gly
195 200 205

Gly Gly Tyr Met Leu Leu Pro Gln Ala Ala Ser Ser Ser Gly Asn Asn
210 215 220

Gly Asn Pro Asn Ala Gly His Met Ser Ser Gly Ser Val Gly Asn Gly
225 230 235 240

Ser Gly Gly Ala Gly Asn Gly Gly Ala Gly Gly Asn Ser Gly Pro Gly
245 250 255

Asn Pro Met Gly Gly Thr Ser Ala Thr Pro Gly His Gly Gly Glu Val
260 265 270

Ile Asp Phe Lys His Leu Phe Glu Glu Leu Cys Pro Val Cys Gly Asp
275 280 285

Lys Val Ser Gly Tyr His Tyr Gly Leu Leu Thr Cys Glu Ser Cys Lys
290 295 300

Gly Phe Phe Lys Arg Thr Val Gln Asn Lys Lys Val Tyr Thr Cys Val
305 310 315 320

Ala Glu Arg Ser Cys His Ile Asp Lys Thr Gln Arg Lys Arg Cys Pro
325 330 335

Tyr Cys Arg Phe Gln Lys Cys Leu Glu Val Gly Met Lys Leu Glu Ala
340 345 350

Val Arg Ala Asp Arg Met Arg Gly Gly Arg Asn Lys Phe Gly Pro Met
355 360 365

Tyr Lys Arg Asp Arg Ala Arg Lys Leu Gln Val Met Arg Gln Arg Gln
370 375 380

Leu Ala Leu Gln Ala Leu Arg Asn Ser Met Gly Pro Asp Ile Lys Pro
385 390 395 400

Thr Pro Ile Ser Pro Gly Tyr Gln Gln Ala Tyr Pro Asn Met Asn Ile
405 410 415

Lys Gln Glu Ile Gln Ile Pro Gln Val Ser Ser Leu Thr Gln Ser Pro
420 425 430

Asp Ser Ser Pro Ser Pro Ile Ala Ile Ala Leu Gly Gln Val Asn Ala
435 440 445

Ser Thr Gly Gly Val Ile Ala Thr Pro Met Asn Ala Gly Thr Gly Gly
450 455 460

```
Ser Gly Gly Gly Gly Leu Asn Gly Pro Ser Ser Val Gly Asn Gly Asn
465                 470                 475                 480

Ser Ser Asn Gly Ser Ser Asn Gly Asn Asn Asn Ser Ser Thr Gly Asn
                485                 490                 495

Gly Thr Ser Gly Gly Gly Gly Gly Asn Asn Ala Gly Gly Gly Gly Gly
            500                 505                 510

Gly Thr Asn Ser Asn Asp Gly Leu His Arg Asn Gly Gly Asn Asp Ser
        515                 520                 525

Ser Ser Cys His Glu Ala Gly Ile Gly Ser Leu Gln Asn Thr Ala Asp
    530                 535                 540

Ser Lys Leu Cys Phe Asp Ser Gly Thr His Pro Ser Ser Thr Ala Asp
545                 550                 555                 560

Ala Leu Ile Glu Pro Leu Arg Val Ser Pro Met Ile Arg Glu Phe Val
                565                 570                 575

Gln Ser Ile Asp Asp Arg Glu Trp Gln Thr Gln Leu Phe Ala Leu Leu
            580                 585                 590

Gln Lys Gln Thr Tyr Asn Gln Val Glu Val Asp Leu Phe Glu Leu Leu
        595                 600                 605

Met Cys Lys Val Leu Asp Gln Asn Leu Phe Ser Gln Val Asp Trp Ala
    610                 615                 620

Arg Asn Thr Val Phe Phe Lys Asp Leu Lys Val Asp Asp Gln Met Lys
625                 630                 635                 640

Leu Leu Gln His Ser Trp Ser Asp Met Leu Val Leu Asp His Leu His
                645                 650                 655

His Arg Ile His Asn Gly Leu Pro Asp Glu Thr Gln Leu Asn Asn Gly
            660                 665                 670

Gln Val Phe Asn Leu Met Ser Leu Gly Leu Leu Gly Val Pro Gln Pro
        675                 680                 685

Gly Asp Tyr Phe Asn Glu Leu Gln Asn Lys Leu Gln Asp Leu Lys Phe
    690                 695                 700

Asp Met Gly Asp Tyr Val Cys Met Lys Phe Leu Ile Leu Leu Asn Pro
705                 710                 715                 720

Ser Val Arg Gly Ile Val Asn Arg Lys Thr Val Ser Glu Gly His Asp
                725                 730                 735

Asn Val Gln Ala Ala Leu Leu Asp Tyr Thr Leu Thr Cys Tyr Pro Ser
            740                 745                 750

Val Asn Asp Lys Phe Arg Gly Leu Val Asn Ile Leu Pro Glu Ile His
        755                 760                 765

Ala Met Ala Val Arg Gly Glu Asp His Leu Ile Thr Cys Thr Pro Ser
    770                 775                 780
```

Thr Val Pro Ala Val Arg Pro Pro Lys Arg Cys Ser Trp Arg Cys Cys
785 790 795 800

Thr Pro Ser Ala Arg Asp Arg Gly Arg Glu Asn Val Thr Arg Asn Thr
805 810 815

<210> 33
<211> 5970
<212> DNA
<213> Drosophila melanogaster
<400> 33
acttactagt gaaaaacatg ataataaaca acttgccaaa aaaaatccaa tgaaattgac 60

acttatgtta aaaaaatagg tgagattgta accgttgatg tacacttacg aagtacgtaa 120

caagttcatg aactgatttc gtgagcaggt ctctccataa tcgccgtatc tgtgggatcg 180

cgcgctcctg ctcgcactcg ctgggtggat ggcagcacat gttcgaagtg cgagagagtg 240

caaagcggag agcgccgacg tcgacgccga aaaaactgaa caagatccgc cgcgaatgtt 300

gattttcctt tcattgacta actgccactc gcagcgcgca gatcgtcggc tccgcttgtt 360

ccgttccgtt cgtttcgttt cgtttcgttc gatctacttc gagtcgcgag ttttaagcag 420

tgtagtgagt gccccgtgaa aaggataacc caaaaagtga tttctactat tttccaatag 480

tttttatcag tgtgaagaaa acatgtaaac ttggctcaaa aagggcttta aaagatacaa 540

agcttcaatg cgaaggataa aataatatcg caccagtgct tcaaaaacca aaactatgcc 600

taaggctgga aatttaaatt aaaatttttt taataaatat tccaaaaata ttgcccctga 660

aaagtgttga taaaccccca accgagcaaa atgttaatgt ccgcggacag ttcagatagc 720

gccaagactt ctgtgatctg cagcacggtg agtgccagca tgctagcacc accagctcca 780

gaacagccca gcaccacagc accacccatt ttgggggtaa caggtcgatc tcacctggaa 840

aatgccctga aactaccgcc aaacacaagt gtttcggctt actaccagca caacagcaag 900

ctgggcatgg gccagaatta caatccggaa ttcaggagcc tggtagcacc tgtcacagat 960

ctggatactg tgccacccac aggtgtgacc atggcgagtt cttcgaattc tcccaactcc 1020

tccgtcaagc tgccccacag cggcgtgatc tttgtcagca aatcgagtgc cgtcagcacc 1080

accgatggtc ccactgcagt gttgcaacag cagcagccgc agcagcaaat gccccagcac 1140

ttcgagtccc tgccccacca ccacccccag caggaacacc agccacagca gcagcagcaa 1200

caacatcacc ttcagcacca cccacatcca catgtgatgt atccgcacgg atatcagcag 1260

gccaatctgc accactcggg tggtattgct gtggttccgg cggattcgcg tccccagact 1320

cccgagtaca tcaagtccta cccagttatg gatacaactg tggctagttc ggtaaagggg 1380

gaaccagaac tcaacataga attcgatggc accacagtgc tgtgccgcgt ttgcggggat 1440

aaggcctccg gtttccatta cggcgtgcat tcctgcgagg gttgcaaggg attcttccgc 1500

cgctccatcc agcaaaagat ccagtatcgc ccgtgcacca agaatcagca gtgcagcatt 1560

ctgcgcatca atcgcaatcg ttgtcaatat tgccgcctga aaaagtgcat tgccgtgggc 1620

atgagtcgcg atgctgtgcg ttttggacgc gtgccgaagc gcgaaaaggc gcgtatcctg 1680

gcggccatgc aacagagcac ccagaatcgc ggccagcagc gagccctcgc caccgagctg 1740

gatgaccagc cacgcctcct cgccgccgtg ctgcgcgccc acctcgagac ctgtgagttc 1800

accaaggaga aggtctcggc gatgcggcag cgggcgcggg attgcccctc ctactccatg 1860

cccacacttc tggcctgtcc gctgaacccc gcccctgaac tgcaatcgga gcaggagttc 1920

tcgcagcgtt tcgcccacgt aattcgcggc gtgatcgact ttgccggcat gattcccggc 1980

ttccagctgc tcacccagga cgataagttc acgctcctga aggcgggact cttcgacgcc 2040

ctgtttgtgc gcctgatctg catgtttgac tcgtcgataa actcaatcat ctgtctaaat 2100

ggccaggtga tgcgacggga tgcgatccag aacggagcca atgcccgctt cctggtggac 2160

tccaccttca atttcgcgga gcgcatgaac tcgatgaacc tgacagatgc cgagataggc 2220

ctgttctgcg ccatcgttct gattacgccg gatcgccccg gtttgcgcaa cctggagctg 2280

atcgagaaga tgtactcgcg actcaagggc tgcctgcagt acattgtcgc ccagaatagg 2340

cccgatcagc ccgagttcct ggccaagttg ctggagacga tgcccgatct gcgcaccctg 2400

agcaccctgc acaccgagaa actggtagtt ttccgcaccg agcacaagga gctgctgcgc 2460

cagcagatgt ggtccatgga ggacggcaac aacagcgatg gccagcagaa caagtcgccc 2520

tcgggcagct gggcggatgc catggacgtg gaggcggcca agagtccgct tggctcggta 2580

tcgagcactg agtccgccga cctggactac ggcagtccga gcagttcgca gccacagggc 2640

gtgtctctgc cctcgccgcc tcagcaacag ccctcggctc tggccagctc ggctcctctg 2700

ctggcggcca ccctctccgg aggatgtccc ctgcgcaacc gggccaattc cggctccagc 2760

ggtgactccg gagcagctga gatggatatc gttggctcgc acgcacatct cacccagaac 2820

gggctgacaa tcacgccgat tgtgcgacac cagcagcagc aacaacagca gcagcagatc 2880

ggaatactca ataatgcgca ttcccgcaac ttgaatgggg gacacgcgat gtgccagcaa 2940

cagcagcagc acccacaact gcaccaccac ttgacagccg gagctgcccg ctacagaaag 3000

ctagattcgc ccacggattc gggcattgag tcgggcaacg agaagaacga gtgcaaggcg 3060

gtgagttcgg ggggaagttc ctcgtgctcc agtccgcgtt ccagtgtgga tgatgcgctg 3120

gactgcagcg atgccgccgc caatcacaat caggtggtgc agcatccgca gctgagtgtg 3180

gtgtccgtgt caccagttcg ctcgccccag ccctccacca gcagccatct gaagcgacag 3240

attgtggagg atatgcccgt gctgaagcgc gtgctgcagg ctccccctct gtacgatacc 3300

aactcgctga tggacgaggc ctacaagccg cacaagaaat tccgggccct gcggcatcgc 3360

gagttcgaga ccgccgaggc ggatgccagc agttccactt ccggctcgaa cagcctgagt 3420

gccggcagtc cgcggcagag cccagtcccg aacagtgtgg ccacgccccc gccatcggcg 3480

gccagcgccg ccgcaggtaa tcccgcccag agccagctgc acatgcacct gacccgcagc 3540

agccccaagg cctcgatggc cagctcgcac tcggtgctgg ccaagtctct catggccgag 3600

ccgcgcatga cgcccgagca gatgaagcgc agcgatatta tccaaaacta cttgaagcgc 3660

gagaacagca cagcagccag cagcaccacc aatggcgtgg gcaaccgcag tcccagcagc 3720

agctccacac cgccgccgtc ggcggtccag aatcagcagc gttggggcag cagctcggtg 3780

atcaccacca cctgccagca gcgccagcag tccgtgtcgc cgcacagcaa cggttccagc 3840

tccagttcga gctctagctc cagctccagt tcgtcatcct cctccacatc ctccaactgc 3900

agctccagct cggccagcag ctgccagtat ttccagtcgc cgcactccac cagcaacggc 3960

accagtgcac cggcgagctc cagttcggga tcgaacagcg ccacgccccc gctggaactg 4020

caggtggaca ttgctgactc ggcgcagcct ctcaatttgt ccaagaaatc gcccacgccg 4080

ccgcccagca agctgcacgc tctggtggcc gccgccaatg ccgttcaaag gtatcccaca 4140

ttgtccgccg acgtcacagt gacagcctcc aatggcgggt cctccgtcgg cggcggcgag 4200

tccggccgcc agcagcagtc cgccggcgag tgtgggctcc cccaatccgg gcctgagcgc 4260

cgccgtgcac aaggtaatgc tggaggcgta agagcgggag gaggtaggtg gttttacgcg 4320

gagaagtggg agagacagag actgggagtg gcagttcagc gaagcaggaa gcaggatcac 4380

ttggagcggc gggagttgaa ttaaattatt ttaccattta attgagacgt gtacaaagtt 4440

tgaaagcaaa accaacatgc atgcaattta aaactaatat ttaaagcaac aacaaacaaa 4500

acaactacaa gttattaatt taaaaaacaa acaaacaaac aaacaacaaa aaacccaagc 4560

ttgaatggta ttacaaaaga aaaagaaaaa cagaaaaaat ataaatatat tttagcagtt 4620

aaactttaac gtagcaagaa accaacaaac ccaaggcagc gctctgattt cgcattaact 4680

tttcttcagc tgctaccgaa aacgcccctc acctcccccc cacccaaccc ttcctccaca 4740

caccaaccgt ctttcgaccc ctgattgttt tataagtttt aagctcttgt tgtacatatt 4800

aattacgttt attggtaact atgtttagcg ctttagttgt agttggagca aaactacttt 4860

gctttttgg atgttttttg aaaaaactgc aaattattat tattaaattt ttaaatacct 4920

aaaaacaaaa caatgtgtgt gaaatttttt attgtgcgat ctccaagcag aatgaagtgc 4980

agtttgcaac aaattttaac tacgattaag ttgataacga ttcatttttt atgaatttaa 5040

ctaattttat gaatttgtta tagttttcca cccttctata gatctttcta tctgatcatc 5100

tagctacccg tattcctgat ttctcctttg gcacaaagct cttctctatg ctaaagaatc 5160

aagtggaata aatattgttt tctaatttta aaactaccac aaaaatacga ttaaaatata 5220

cacgaagtaa ttgaaaatca aacaaaatgc ttaaagtttt agcagcaagc agtaaaacga 5280

cgatgaagaa gagaaaccca acgttaaata tatctgttgt gtacatagtt aaatgttaaa 5340

ttaaacacaa aaacatattt aaagtacata taaatacaca taattattaa tgaagaaacc 5400

tatgcttaaa agattcaatg tttgattggc atcttagaaa accaagcgaa aaatacaaaa 5460

aaaaatcaac aaacaaaaat tatgatatat tatttaaaag taaagtatac atttacatta 5520

cagaaaaaca aaagagaaaa cttgcggtag caacaaaact attatattaa ttacatttta 5580

attatgctgt actattatga ttattaatta ttatgattaa ttaattacga tttttatgct 5640

tagacaaacc aacaaaaaac aaatatgcaa aaaccattaa aaaaaaaaac aaaaaacaag 5700

caaaaaatta cactggcgca gaaatttgta ttgtcaaatc aagaaaaaat ttgtttaaaa 5760

atttcaaaag aaagcctctt aagtttttc atttcaattt ccttttcagt tgaacacttt 5820

ttcttaaaat gttcagtttt accgttactt tgctttggag tggtaagatt ttgggtttca 5880

cttgatttca ttttcgtttt gactatcgtg ctgtaaaaat atttactaaa ttatatcgaa 5940

agtatttata tcataattaa ataaacaaga 5970

<210> 34
<211> 1237
<212> PRT
<213> Drosophila melanogaster
<400> 34

Met Leu Met Ser Ala Asp Ser Ser Asp Ser Ala Lys Thr Ser Val Ile
1 5 10 15

Cys Ser Thr Val Ser Ala Ser Met Leu Ala Pro Pro Ala Pro Glu Gln
20 25 30

Pro Ser Thr Thr Ala Pro Pro Ile Leu Gly Val Thr Gly Arg Ser His
35 40 45

Leu Glu Asn Ala Leu Lys Leu Pro Pro Asn Thr Ser Val Ser Ala Tyr
50 55 60

Tyr Gln His Asn Ser Lys Leu Gly Met Gly Gln Asn Tyr Asn Pro Glu
65 70 75 80

Phe Arg Ser Leu Val Ala Pro Val Thr Asp Leu Asp Thr Val Pro Pro
85 90 95

Thr Gly Val Thr Met Ala Ser Ser Ser Asn Ser Pro Asn Ser Ser Val
100 105 110

Lys Leu Pro His Ser Gly Val Ile Phe Val Ser Lys Ser Ser Ala Val
115 120 125

```
Ser Thr Thr Asp Gly Pro Thr Ala Val Leu Gln Gln Gln Gln Pro Gln
    130                 135                 140

Gln Gln Met Pro Gln His Phe Glu Ser Leu Pro His His His Pro Gln
145                 150                 155                 160

Gln Glu His Gln Pro Gln Gln Gln Gln Gln Gln His His Leu Gln His
                165                 170                 175

His Pro His Pro His Val Met Tyr Pro His Gly Tyr Gln Gln Ala Asn
            180                 185                 190

Leu His His Ser Gly Gly Ile Ala Val Val Pro Ala Asp Ser Arg Pro
        195                 200                 205

Gln Thr Pro Glu Tyr Ile Lys Ser Tyr Pro Val Met Asp Thr Thr Val
    210                 215                 220

Ala Ser Ser Val Lys Gly Glu Pro Glu Leu Asn Ile Glu Phe Asp Gly
225                 230                 235                 240

Thr Thr Val Leu Cys Arg Val Cys Gly Asp Lys Ala Ser Gly Phe His
                245                 250                 255

Tyr Gly Val His Ser Cys Glu Gly Cys Lys Gly Phe Phe Arg Arg Ser
            260                 265                 270

Ile Gln Gln Lys Ile Gln Tyr Arg Pro Cys Thr Lys Asn Gln Gln Cys
        275                 280                 285

Ser Ile Leu Arg Ile Asn Arg Asn Arg Cys Gln Tyr Cys Arg Leu Lys
    290                 295                 300

Lys Cys Ile Ala Val Gly Met Ser Arg Asp Ala Val Arg Phe Gly Arg
305                 310                 315                 320

Val Pro Lys Arg Glu Lys Ala Arg Ile Leu Ala Ala Met Gln Gln Ser
                325                 330                 335

Thr Gln Asn Arg Gly Gln Gln Arg Ala Leu Ala Thr Glu Leu Asp Asp
            340                 345                 350

Gln Pro Arg Leu Leu Ala Ala Val Leu Arg Ala His Leu Glu Thr Cys
        355                 360                 365

Glu Phe Thr Lys Glu Lys Val Ser Ala Met Arg Gln Arg Ala Arg Asp
    370                 375                 380

Cys Pro Ser Tyr Ser Met Pro Thr Leu Leu Ala Cys Pro Leu Asn Pro
385                 390                 395                 400

Ala Pro Glu Leu Gln Ser Glu Gln Glu Phe Ser Gln Arg Phe Ala His
                405                 410                 415

Val Ile Arg Gly Val Ile Asp Phe Ala Gly Met Ile Pro Gly Phe Gln
            420                 425                 430

Leu Leu Thr Gln Asp Asp Lys Phe Thr Leu Leu Lys Ala Gly Leu Phe
        435                 440                 445
```

```
Asp Ala Leu Phe Val Arg Leu Ile Cys Met Phe Asp Ser Ser Ile Asn
    450                 455                 460

Ser Ile Ile Cys Leu Asn Gly Gln Val Met Arg Arg Asp Ala Ile Gln
465                 470                 475                 480

Asn Gly Ala Asn Ala Arg Phe Leu Val Asp Ser Thr Phe Asn Phe Ala
                485                 490                 495

Glu Arg Met Asn Ser Met Asn Leu Thr Asp Ala Glu Ile Gly Leu Phe
            500                 505                 510

Cys Ala Ile Val Leu Ile Thr Pro Asp Arg Pro Gly Leu Arg Asn Leu
        515                 520                 525

Glu Leu Ile Glu Lys Met Tyr Ser Arg Leu Lys Gly Cys Leu Gln Tyr
    530                 535                 540

Ile Val Ala Gln Asn Arg Pro Asp Gln Pro Glu Phe Leu Ala Lys Leu
545                 550                 555                 560

Leu Glu Thr Met Pro Asp Leu Arg Thr Leu Ser Thr Leu His Thr Glu
                565                 570                 575

Lys Leu Val Val Phe Arg Thr Glu His Lys Glu Leu Leu Arg Gln Gln
            580                 585                 590

Met Trp Ser Met Glu Asp Gly Asn Asn Ser Asp Gly Gln Gln Asn Lys
        595                 600                 605

Ser Pro Ser Gly Ser Trp Ala Asp Ala Met Asp Val Glu Ala Ala Lys
    610                 615                 620

Ser Pro Leu Gly Ser Val Ser Ser Thr Glu Ser Ala Asp Leu Asp Tyr
625                 630                 635                 640

Gly Ser Pro Ser Ser Ser Gln Pro Gln Gly Val Ser Leu Pro Ser Pro
                645                 650                 655

Pro Gln Gln Gln Pro Ser Ala Leu Ala Ser Ser Ala Pro Leu Leu Ala
            660                 665                 670

Ala Thr Leu Ser Gly Gly Cys Pro Leu Arg Asn Arg Ala Asn Ser Gly
        675                 680                 685

Ser Ser Gly Asp Ser Gly Ala Ala Glu Met Asp Ile Val Gly Ser His
    690                 695                 700

Ala His Leu Thr Gln Asn Gly Leu Thr Ile Thr Pro Ile Val Arg His
705                 710                 715                 720

Gln Gln Gln Gln Gln Gln Gln Gln Gln Ile Gly Ile Leu Asn Asn Ala
                725                 730                 735

His Ser Arg Asn Leu Asn Gly Gly His Ala Met Cys Gln Gln Gln Gln
            740                 745                 750

Gln His Pro Gln Leu His His His Leu Thr Ala Gly Ala Ala Arg Tyr
        755                 760                 765
```

```
Arg Lys Leu Asp Ser Pro Thr Asp Ser Gly Ile Glu Ser Gly Asn Glu
    770                 775                 780

Lys Asn Glu Cys Lys Ala Val Ser Ser Gly Gly Ser Ser Ser Cys Ser
785                 790                 795                 800

Ser Pro Arg Ser Ser Val Asp Asp Ala Leu Asp Cys Ser Asp Ala Ala
                805                 810                 815

Ala Asn His Asn Gln Val Val Gln His Pro Gln Leu Ser Val Val Ser
            820                 825                 830

Val Ser Pro Val Arg Ser Pro Gln Pro Ser Thr Ser Ser His Leu Lys
        835                 840                 845

Arg Gln Ile Val Glu Asp Met Pro Val Leu Lys Arg Val Leu Gln Ala
    850                 855                 860

Pro Pro Leu Tyr Asp Thr Asn Ser Leu Met Asp Glu Ala Tyr Lys Pro
865                 870                 875                 880

His Lys Lys Phe Arg Ala Leu Arg His Arg Glu Phe Glu Thr Ala Glu
                885                 890                 895

Ala Asp Ala Ser Ser Ser Thr Ser Gly Ser Asn Ser Leu Ser Ala Gly
            900                 905                 910

Ser Pro Arg Gln Ser Pro Val Pro Asn Ser Val Ala Thr Pro Pro Pro
        915                 920                 925

Ser Ala Ala Ser Ala Ala Ala Gly Asn Pro Ala Gln Ser Gln Leu His
    930                 935                 940

Met His Leu Thr Arg Ser Ser Pro Lys Ala Ser Met Ala Ser Ser His
945                 950                 955                 960

Ser Val Leu Ala Lys Ser Leu Met Ala Glu Pro Arg Met Thr Pro Glu
                965                 970                 975

Gln Met Lys Arg Ser Asp Ile Ile Gln Asn Tyr Leu Lys Arg Glu Asn
            980                 985                 990

Ser Thr Ala Ala Ser Ser Thr Thr  Asn Gly Val Gly Asn  Arg Ser Pro
        995                 1000                1005

Ser Ser  Ser Ser Thr Pro Pro  Pro Ser Ala Val Gln  Asn Gln Gln
    1010                1015                1020

Arg Trp  Gly Ser Ser Ser Val  Ile Thr Thr Thr Cys  Gln Gln Arg
    1025                1030                1035

Gln Gln  Ser Val Ser Pro His  Ser Asn Gly Ser Ser  Ser Ser Ser
    1040                1045                1050

Ser Ser  Ser Ser Ser Ser Ser  Ser Ser Ser Ser Ser  Thr Ser Ser
    1055                1060                1065

Asn Cys  Ser Ser Ser Ser Ala  Ser Ser Cys Gln Tyr  Phe Gln Ser
    1070                1075                1080
```

Pro His Ser Thr Ser Asn Gly Thr Ser Ala Pro Ala Ser Ser Ser
1085 1090 1095

Ser Gly Ser Asn Ser Ala Thr Pro Leu Leu Glu Leu Gln Val Asp
1100 1105 1110

Ile Ala Asp Ser Ala Gln Pro Leu Asn Leu Ser Lys Lys Ser Pro
1115 1120 1125

Thr Pro Pro Pro Ser Lys Leu His Ala Leu Val Ala Ala Ala Asn
1130 1135 1140

Ala Val Gln Arg Tyr Pro Thr Leu Ser Ala Asp Val Thr Val Thr
1145 1150 1155

Ala Ser Asn Gly Gly Ser Ser Val Gly Gly Gly Glu Ser Gly Arg
1160 1165 1170

Gln Gln Gln Ser Ala Gly Glu Cys Gly Leu Pro Gln Ser Gly Pro
1175 1180 1185

Glu Arg Arg Arg Ala Gln Gly Asn Ala Gly Gly Val Arg Ala Gly
1190 1195 1200

Gly Gly Arg Trp Phe Tyr Ala Glu Lys Trp Glu Arg Gln Arg Leu
1205 1210 1215

Gly Val Ala Val Gln Arg Ser Arg Lys Gln Asp His Leu Glu Arg
1220 1225 1230

Arg Glu Leu Asn
1235

<210> 35
<211> 25
<212> DNA
<213> Drosophila melanogaster
<220>
<221> misc_feature
<222> (1)..(25)
<223> n is a, c, g, or t
<220>
<221> misc_feature
<222> (1)..(25)
<223> y is t/u or g
<400> 35
cgggatcctg yaarggntty ttyaa 25

<210> 36
<211> 25
<212> DNA
<213> Drosophila melanogaster
<220>
<221> misc_feature
<222> (1)..(25)
<223> y is t/u or g
<220>
<221> misc_feature
<222> (1)..(25)

<223> h is a, c, or t/u
<400> 36
cgggatcctg ycayathgay aarac 25

<210> 37
<211> 30
<212> DNA
<213> Drosophila melanogaster
<220>
<221> misc_feature
<222> (1)..(30)
<223> n is a, c, g, or t
<220>
<221> misc_feature
<222> (1)..(30)
<223> y is t/u or g
<220>
<221> misc_feature
<222> (1)..(30)
<223> r is g or a
<400> 37
aargtngayg aycaratgaa raagcttggg 30

<210> 38
<211> 27
<212> DNA
<213> Drosophila melanogaster
<220>
<221> misc_feature
<222> (1)..(27)
<223> n is a, c, g, or t
<220>
<221> misc_feature
<222> (1)..(27)
<223> y is t/u or g
<220>
<221> misc_feature
<222> (1)..(27)
<223> r is g or a
<400> 38
aayacngtnt tyttyaaraa gcttggg 27
<210> 39
<211> 27
<212> DNA
<213> Heliothis virescens
<400> 39
aatgtctgca ctgcgatttg tcgctgt 27

<210> 40
<211> 28
<212> DNA
<213> Heliothis virescens
<400> 40
acagatccct caagtgtcgt ctctcacg 28

<210> 41
<211> 27

<212> DNA
<213> Heliothis virescens
<400> 41
attgtaagtc tggctctgca agagccc 27

<210> 42
<211> 28
<212> DNA
<213> Heliothis virescens
<400> 42
cgtgagagac gacacttgag ggatctgt 28

<210> 43
<211> 462
<212> PRT
<213> Bombyx mori
<400> 43

Met Ser Ser Val Ala Lys Lys Asp Lys Arg Thr Met Ser Val Thr Ala
1 5 10 15

Leu Ile Asn Arg Ala Trp Pro Met Thr Pro Ser Pro Gln Gln Gln Gln
20 25 30

Gln Met Val Pro Ser Thr Gln His Ser Asn Phe Leu His Ala Met Ala
35 40 45

Thr Pro Ser Thr Thr Pro Asn Val Glu Leu Asp Ile Gln Trp Leu Asn
50 55 60

Ile Glu Ser Gly Phe Met Ser Pro Met Ser Pro Pro Glu Met Lys Pro
65 70 75 80

Asp Thr Ala Met Leu Asp Gly Phe Arg Asp Asp Ser Thr Pro Pro Pro
85 90 95

Pro Phe Lys Asn Tyr Pro Pro Asn His Pro Leu Ser Gly Ser Lys His
100 105 110

Leu Cys Ser Ile Cys Gly Asp Arg Ala Ser Gly Lys His Tyr Gly Val
115 120 125

Tyr Ser Cys Glu Gly Cys Lys Gly Phe Phe Lys Arg Thr Val Arg Lys
130 135 140

Asp Leu Thr Tyr Ala Cys Arg Glu Asp Lys Asn Cys Ile Ile Asp Lys
145 150 155 160

Arg Gln Arg Asn Arg Cys Gln Tyr Cys Arg Tyr Gln Lys Cys Leu Ala
165 170 175

Cys Gly Met Lys Arg Glu Ala Val Gln Glu Glu Arg Gln Arg Ala Ala
180 185 190

Arg Arg Thr Glu Asp Ala His Pro Ser Ser Ser Val Gln Glu Leu Ser
195 200 205

Ile Glu Arg Leu Leu Glu Leu Glu Ala Leu Val Ala Asp Ser Ala Glu
210 215 220

Glu Leu Gln Ile Leu Arg Val Gly Pro Glu Ser Gly Val Pro Ala Lys
225 230 235 240

Tyr Arg Ala Pro Val Ser Ser Leu Cys Gln Ile Gly Asn Lys Gln Ile
245 250 255

Ala Ala Leu Ile Val Trp Ala Arg Asp Ile Pro His Phe Gly Gln Leu
260 265 270

Glu Ile Asp Asp Gln Ile Leu Leu Ile Lys Gly Ser Trp Asn Glu Leu
275 280 285

Leu Leu Phe Ala Ile Ala Trp Arg Ser Met Glu Phe Leu Asn Asp Glu
290 295 300

Arg Glu Asn Val Asp Ser Arg Asn Thr Ala Pro Pro Gln Leu Ile Cys
305 310 315 320

Leu Met Pro Gly Met Thr Leu His Arg Asn Ser Ala Leu Gln Ala Gly
325 330 335

Val Gly Gln Ile Phe Asp Arg Val Leu Ser Glu Leu Ser Leu Lys Met
340 345 350

Arg Ser Leu Arg Met Asp Gln Ala Glu Cys Val Ala Leu Lys Ala Ile
355 360 365

Ile Leu Leu Asn Pro Asp Val Lys Gly Leu Lys Asn Lys Gln Glu Val
370 375 380

Asp Val Leu Arg Glu Lys Met Phe Leu Cys Leu Asp Glu Tyr Cys Arg
385 390 395 400

Arg Ser Arg Gly Gly Glu Glu Gly Arg Phe Ala Ala Leu Leu Leu Arg
405 410 415

Leu Pro Ala Leu Arg Ser Ile Ser Leu Lys Ser Phe Glu His Leu Tyr
420 425 430

Leu Phe His Leu Val Ala Glu Gly Ser Val Ser Ser Tyr Ile Arg Asp
435 440 445

Ala Leu Cys Asn His Ala Pro Pro Ile Asp Thr Asn Ile Met
450 455 460

<210> 44
<211> 461
<212> PRT
<213> Manduca sexta
<400> 44

Met Ser Ser Val Ala Lys Lys Asp Lys Arg Thr Met Ser Val Thr Ala
1 5 10 15

Leu Ile Asn Arg Ala Trp Pro Leu Thr Pro Ala Pro His Gln Gln Gln
20 25 30

Ser Met Pro Ser Ser Gln Pro Ser Asn Phe Leu Gln Pro Leu Ala Thr
35 40 45

```
Pro Ser Thr Thr Pro Ser Val Glu Leu Asp Ile Gln Trp Leu Asn Ile
    50                  55                  60

Glu Pro Gly Phe Met Ser Pro Met Ser Pro Pro Glu Met Lys Pro Asp
65                  70                  75                  80

Thr Ala Met Leu Asp Gly Leu Arg Asp Asp Ser Thr Pro Pro Pro Ala
                85                  90                  95

Phe Lys Asn Tyr Pro Pro Asn His Pro Leu Ser Gly Ser Lys His Leu
            100                 105                 110

Cys Ser Ile Cys Gly Asp Arg Ala Ser Gly Lys His Tyr Gly Val Tyr
        115                 120                 125

Ser Cys Glu Gly Cys Lys Gly Phe Phe Lys Arg Thr Val Arg Lys Asp
    130                 135                 140

Leu Thr Tyr Ala Cys Arg Glu Asp Arg Asn Cys Ile Ile Asp Lys Arg
145                 150                 155                 160

Gln Arg Asn Arg Cys Gln Tyr Cys Arg Tyr Gln Lys Cys Leu Ala Cys
                165                 170                 175

Gly Met Lys Arg Glu Ala Val Gln Glu Glu Arg Gln Arg Ala Ala Arg
            180                 185                 190

Gly Thr Glu Asp Ala His Pro Ser Ser Ser Val Gln Glu Leu Ser Ile
        195                 200                 205

Glu Arg Leu Leu Glu Ile Glu Ser Leu Val Ala Asp Pro Pro Glu Glu
    210                 215                 220

Phe Gln Phe Leu Arg Val Gly Pro Glu Ser Gly Val Pro Ala Lys Tyr
225                 230                 235                 240

Arg Ala Pro Val Ser Ser Leu Cys Gln Ile Gly Asn Lys Gln Ile Ala
                245                 250                 255

Ala Leu Val Val Trp Ala Arg Asp Ile Pro His Phe Gly Gln Leu Glu
            260                 265                 270

Leu Glu Asp Gln Ile Leu Leu Ile Lys Asn Ser Trp Asn Glu Leu Leu
        275                 280                 285

Leu Phe Ala Ile Ala Trp Arg Ser Met Glu Tyr Leu Thr Asp Glu Arg
    290                 295                 300

Glu Asn Val Asp Ser Arg Ser Thr Ala Pro Pro Gln Leu Met Cys Leu
305                 310                 315                 320

Met Pro Gly Met Thr Leu His Arg Asn Ser Ala Leu Gln Ala Gly Val
                325                 330                 335

Gly Gln Ile Phe Asp Arg Val Leu Ser Glu Leu Ser Leu Lys Met Arg
            340                 345                 350

Thr Leu Arg Met Asp Gln Ala Glu Tyr Val Ala Leu Lys Ala Ile Ile
        355                 360                 365
```

Leu Leu Asn Pro Asp Val Lys Gly Leu Lys Asn Lys Pro Glu Val Val
370 375 380

Val Leu Arg Glu Lys Met Phe Ser Cys Leu Asp Glu Tyr Val Arg Arg
385 390 395 400

Ser Arg Cys Ala Glu Glu Gly Arg Phe Ala Ala Leu Leu Leu Arg Leu
405 410 415

Pro Ala Leu Arg Ser Ile Ser Leu Lys Cys Phe Glu His Leu Tyr Phe
420 425 430

Phe His Leu Val Ala Asp Thr Ser Ile Ala Ser Tyr Ile His Asp Ala
435 440 445

Leu Arg Asn His Ala Pro Ser Ile Asp Thr Ser Ile Leu
450 455 460

<210> 45
<211> 472
<212> PRT
<213> Choristoneura fumiferana
<400> 45

Met Ser Ser Val Ala Lys Lys Asp Lys Pro Thr Met Ser Val Thr Ala
1 5 10 15

Leu Ile Asn Trp Ala Arg Pro Ala Pro Pro Gly Pro Pro Gln Pro Gln
20 25 30

Ser Ala Ser Pro Ala Pro Ala Ala Met Leu Gln Gln Leu Pro Thr Gln
35 40 45

Ser Met Gln Ser Leu Asn His Ile Pro Thr Val Asp Cys Ser Leu Asp
50 55 60

Met Gln Trp Leu Asn Leu Glu Pro Gly Phe Met Ser Pro Met Ser Pro
65 70 75 80

Pro Glu Met Lys Pro Asp Thr Ala Met Leu Asp Gly Leu Arg Asp Asp
85 90 95

Ala Thr Ser Pro Pro Asn Phe Lys Asn Tyr Pro Pro Asn His Pro Leu
100 105 110

Ser Gly Ser Lys His Leu Cys Ser Ile Cys Gly Asp Arg Ala Ser Gly
115 120 125

Lys His Tyr Gly Val Tyr Ser Cys Glu Gly Cys Lys Gly Phe Phe Lys
130 135 140

Arg Thr Val Arg Lys Asp Leu Ser Tyr Ala Cys Arg Glu Glu Arg Asn
145 150 155 160

Cys Ile Ile Asp Lys Arg Gln Arg Asn Arg Cys Gln Tyr Cys Arg Tyr
165 170 175

Gln Lys Cys Leu Ala Cys Gly Met Lys Arg Glu Ala Val Gln Glu Glu
180 185 190

```
Arg Gln Arg Asn Ala Arg Gly Ala Glu Asp Ala His Pro Ser Ser Ser
        195                 200                 205

Val Gln Val Ser Asp Glu Leu Ser Ile Glu Arg Leu Thr Glu Met Glu
    210                 215                 220

Ser Leu Val Ala Asp Pro Ser Glu Glu Phe Gln Phe Leu Arg Val Gly
225                 230                 235                 240

Pro Asp Ser Asn Val Pro Pro Arg Tyr Arg Ala Pro Val Ser Ser Leu
                245                 250                 255

Cys Gln Ile Gly Asn Lys Gln Ile Ala Ala Leu Val Val Trp Ala Arg
            260                 265                 270

Asp Ile Pro His Phe Gly Gln Leu Glu Leu Asp Asp Gln Val Val Leu
        275                 280                 285

Ile Lys Ala Ser Trp Asn Glu Leu Leu Leu Phe Ala Ile Ala Trp Arg
    290                 295                 300

Ser Met Glu Tyr Leu Glu Asp Glu Arg Glu Asn Gly Asp Gly Thr Arg
305                 310                 315                 320

Ser Thr Thr Gln Pro Gln Leu Met Cys Leu Met Pro Gly Met Thr Leu
                325                 330                 335

His Arg Asn Ser Ala Gln Gln Ala Gly Val Gly Ala Ile Phe Asp Arg
            340                 345                 350

Val Leu Ser Glu Leu Ser Leu Lys Met Arg Thr Leu Arg Met Asp Gln
        355                 360                 365

Ala Glu Tyr Val Ala Leu Lys Ala Ile Val Leu Leu Asn Pro Asp Val
    370                 375                 380

Lys Gly Leu Lys Asn Arg Gln Glu Val Asp Val Leu Arg Glu Lys Met
385                 390                 395                 400

Phe Ser Cys Leu Asp Asp Tyr Cys Arg Arg Ser Arg Ser Asn Glu Glu
                405                 410                 415

Gly Arg Phe Ala Ser Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile
            420                 425                 430

Ser Leu Lys Ser Phe Glu His Leu Tyr Phe Phe His Leu Val Ala Glu
        435                 440                 445

Gly Ser Ile Ser Gly Tyr Ile Arg Glu Ala Leu Arg Asn His Ala Pro
    450                 455                 460

Pro Ile Asp Val Asn Ala Met Met
465                 470
```

<210> 46
<211> 389
<212> PRT
<213> Locusta migratoria
<400> 46

```
Met Glu Gly Ser Glu Arg Gly Ile Ser Leu Glu Asn Asn Leu Ser Ile
1               5                  10                  15

Ser Ser Met Gly Pro Gln Ser Pro Leu Asp Met Lys Pro Asp Thr Ala
            20                  25                  30

Ser Leu Ile Ser Ser Gly Ser Phe Ser Pro Thr Gly Gly Pro Asn Ser
        35                  40                  45

Pro Gly Ser Phe Thr Ile Gly His Ser Ser Leu Leu Asn Asn Ser Ser
    50                  55                  60

Ser Asn Gln Ala Lys Gly Ser Ser Ser Gln Tyr Pro Pro Asn His Pro
65                  70                  75                  80

Leu Ser Gly Ser Lys His Leu Cys Ser Ile Cys Gly Asp Arg Ala Ser
                85                  90                  95

Gly Lys His Tyr Gly Val Tyr Ser Cys Glu Gly Cys Lys Gly Phe Phe
            100                 105                 110

Lys Arg Thr Val Arg Lys Asp Leu Ser Tyr Ala Cys Arg Glu Asp Lys
        115                 120                 125

Asn Cys Ile Ile Asp Lys Arg Gln Arg Asn Arg Cys Gln Tyr Cys Arg
    130                 135                 140

Tyr Gln Lys Cys Leu Ala Met Gly Met Lys Arg Glu Ala Val Gln Glu
145                 150                 155                 160

Glu Arg Gln Arg Thr Lys Glu Arg Asp Gln Asn Glu Val Glu Ser Thr
                165                 170                 175

Ser Ser Leu His Thr Asp Met Pro Val Glu Arg Ile Leu Glu Ala Glu
            180                 185                 190

Lys Arg Val Glu Cys Lys Ala Glu Asn Gln Val Glu Tyr Glu Leu Val
        195                 200                 205

Glu Trp Ala Lys His Ile Pro His Phe Thr Ser Leu Pro Leu Glu Asp
    210                 215                 220

Gln Val Leu Leu Leu Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ala
225                 230                 235                 240

Phe Ser His Arg Ser Val Asp Val Lys Asp Gly Ile Val Leu Ala Thr
                245                 250                 255

Gly Leu Thr Val His Arg Asn Ser Ala His Gln Ala Gly Val Gly Thr
            260                 265                 270

Ile Phe Asp Arg Val Leu Thr Glu Leu Val Ala Lys Met Arg Glu Met
        275                 280                 285

Lys Met Asp Lys Thr Glu Leu Gly Cys Leu Arg Ser Val Ile Leu Phe
    290                 295                 300

Asn Pro Glu Val Arg Gly Leu Lys Ser Ala Gln Glu Val Glu Leu Leu
305                 310                 315                 320
```

Arg Glu Lys Val Tyr Ala Ala Leu Glu Glu Tyr Thr Arg Thr Thr His
325 330 335

Pro Asp Glu Pro Gly Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ser
340 345 350

Leu Arg Ser Ile Gly Leu Lys Cys Leu Glu His Leu Phe Phe Phe Arg
355 360 365

Leu Ile Gly Asp Val Pro Ile Asp Thr Phe Leu Met Glu Met Leu Glu
370 375 380

Ser Pro Ser Asp Ser
385

<210> 47
<211> 552
<212> PRT
<213> Chironomus tentans

<400> 47

Met Leu Lys Lys Glu Lys Pro Met Met Thr Val Ala Ala Ile Ile Glu
1 5 10 15

Gln Ala Gln Asn Arg Trp Met Asp His Pro Leu Val Tyr Asn Ser Arg
20 25 30

Ser Leu Gln Phe Gln Gly Ser Tyr Cys Ile Asp Ser Ser Leu Leu Gly
35 40 45

His Met Gly Pro Leu Ser Pro Pro Asp Leu Lys Pro Asp Ile Ser Leu
50 55 60

Leu Asn Cys Asn Asn Asn Asn Asn Asn Thr Asn Asn Asn Asn Ser Asn
65 70 75 80

Ser Ser His Asn Asn Leu Asn His His Asn Thr Ser Pro Leu Pro Val
85 90 95

Leu Gly Ala Asn Thr Phe Ser Pro Ile Gln Ser Leu Asn Asn Asn Gly
100 105 110

Pro Ser Ser Pro Leu Ser Ser Ile Gly Asn Gly Ser Gly Thr Ile Val
115 120 125

Thr Phe Asn Gln Ile Lys Leu Gln Ser Pro Ser Pro Ser Asn Ala Ser
130 135 140

Ser Ser Ser Thr Leu Ser Gly Pro Leu Thr Thr Thr Pro Pro Ala Thr
145 150 155 160

Asn Ala Asn Asn Ile Leu Gly Met Gly Asn Gly Asn Cys Gly Asn Thr
165 170 175

Ala Asn Gly Lys Gln Ser Gln Tyr Pro Pro Asn His Pro Leu Ser Gly
180 185 190

Ser Lys His Leu Cys Ser Ile Cys Gly Asp Arg Ala Ser Gly Lys His
195 200 205

```
Tyr Gly Val Tyr Ser Cys Glu Gly Cys Lys Gly Phe Phe Lys Arg Thr
    210                 215                 220

Val Arg Lys Asp Leu Ser Tyr Ala Cys Arg Glu Glu Arg Asn Cys Val
225                 230                 235                 240

Ile Asp Lys Lys Gln Arg Asn Arg Cys Gln Tyr Cys Arg Tyr Gln Lys
                245                 250                 255

Cys Leu Asn Cys Gly Met Lys Arg Glu Ala Val Gln Glu Glu Arg Gln
            260                 265                 270

Arg Gly Gly Lys Ser Gln Lys Gly Asp Asp Met Ser Ile Ser Ser Thr
        275                 280                 285

Gln Ser Leu Val Asn Asn Gly Pro Gly Arg Asp Ile Thr Val Glu Arg
    290                 295                 300

Leu Met Glu Ala Asp Gln Met Ser Glu Ala Arg Cys Gly Asp Lys Ser
305                 310                 315                 320

Ile Gln Tyr Leu Arg Val Ala Ala Ser Asn Thr Met Ile Pro Pro Glu
                325                 330                 335

Tyr Arg Ala Pro Val Ser Ala Ile Cys Ala Met Val Asn Lys Gln Val
            340                 345                 350

Phe Gln His Met Asp Phe Cys Arg Arg Leu Pro His Phe Thr Lys Leu
        355                 360                 365

Pro Leu Asn Asp Gln Met Tyr Leu Leu Lys Gln Ser Leu Asn Glu Leu
    370                 375                 380

Leu Ile Leu Asn Ile Ala Tyr Met Ser Ile Gln Tyr Val Glu Pro Asp
385                 390                 395                 400

Arg Arg Asn Ala Asp Gly Ser Leu Glu Arg Arg Gln Ile Ser Gln Gln
                405                 410                 415

Met Cys Leu Ser Arg Asn Tyr Thr Leu Gly Arg Asn Met Ala Val Gln
            420                 425                 430

Ala Gly Val Val Gln Ile Phe Asp Arg Ile Leu Ser Glu Leu Ser Val
        435                 440                 445

Lys Met Lys Arg Leu Asp Leu Asp Ala Thr Glu Leu Cys Leu Leu Lys
    450                 455                 460

Ser Ile Val Val Phe Asn Pro Asp Val Arg Thr Leu Asp Asp Arg Lys
465                 470                 475                 480

Ser Ile Asp Leu Leu Arg Ser Arg Ile Tyr Ala Ser Leu Asp Glu Tyr
                485                 490                 495

Cys Arg Gln Lys His Pro Asn Glu Asp Gly Arg Phe Ala Gln Leu Leu
            500                 505                 510

Leu Arg Leu Pro Ala Leu Arg Ser Ile Ser Leu Lys Cys Leu Asp His
        515                 520                 525
```

Leu Phe Tyr Phe Gln Leu Ile Asp Asp Lys Asn Val Glu Asn Ser Val
530 535 540

Ile Glu Glu Phe His Lys Leu Asn
545 550

<210> 48
<211> 508
<212> PRT
<213> Drosophila melanogaster
<400> 48

Met Asp Asn Cys Asp Gln Asp Ala Ser Phe Arg Leu Ser His Ile Lys
1 5 10 15

Glu Glu Val Lys Pro Asp Ile Ser Gln Leu Asn Asp Ser Asn Asn Ser
20 25 30

Ser Phe Ser Pro Lys Ala Glu Ser Pro Val Pro Phe Met Gln Ala Met
35 40 45

Ser Met Val His Val Leu Pro Gly Ser Asn Ser Ala Ser Ser Asn Asn
50 55 60

Asn Ser Ala Gly Asp Ala Gln Met Ala Gln Ala Pro Asn Ser Ala Gly
65 70 75 80

Gly Ser Ala Ala Ala Ala Val Gln Gln Gln Tyr Pro Pro Asn His Pro
85 90 95

Leu Ser Gly Ser Lys His Leu Cys Ser Ile Cys Gly Asp Arg Ala Ser
100 105 110

Gly Lys His Tyr Gly Val Tyr Ser Cys Glu Gly Cys Lys Gly Phe Phe
115 120 125

Lys Arg Thr Val Arg Lys Asp Leu Thr Tyr Ala Cys Arg Glu Asn Arg
130 135 140

Asn Cys Ile Ile Asp Lys Arg Gln Arg Asn Arg Cys Gln Tyr Cys Arg
145 150 155 160

Tyr Gln Lys Cys Leu Thr Cys Gly Met Lys Arg Glu Ala Val Gln Glu
165 170 175

Glu Arg Gln Arg Gly Ala Arg Asn Ala Ala Gly Arg Leu Ser Ala Ser
180 185 190

Gly Gly Gly Ser Ser Gly Pro Gly Ser Val Gly Gly Ser Ser Ser Gln
195 200 205

Gly Gly Gly Gly Gly Gly Gly Val Ser Gly Gly Met Gly Ser Gly Asn
210 215 220

Gly Ser Asp Asp Phe Met Thr Asn Ser Val Ser Arg Asp Phe Ser Ile
225 230 235 240

Glu Arg Ile Ile Glu Ala Glu Gln Arg Ala Glu Thr Gln Cys Gly Asp
245 250 255

Arg Ala Leu Thr Phe Leu Arg Val Gly Pro Tyr Ser Thr Val Gln Pro
260 265 270

Asp Tyr Lys Gly Ala Val Ser Ala Leu Cys Gln Val Val Asn Lys Gln
275 280 285

Leu Phe Gln Met Val Glu Tyr Ala Arg Met Met Pro His Phe Ala Gln
290 295 300

Val Pro Leu Asp Asp Gln Val Ile Leu Leu Lys Ala Ala Trp Ile Glu
305 310 315 320

Leu Leu Ile Ala Asn Val Ala Trp Cys Ser Ile Val Ser Leu Asp Asp
325 330 335

Gly Gly Ala Gly Gly Gly Gly Gly Gly Leu Gly His Asp Gly Ser Phe
340 345 350

Glu Arg Arg Ser Pro Gly Leu Gln Pro Gln Gln Leu Phe Leu Asn Gln
355 360 365

Ser Phe Ser Tyr His Arg Asn Ser Ala Ile Lys Ala Gly Val Ser Ala
370 375 380

Ile Phe Asp Arg Ile Leu Ser Glu Leu Ser Val Lys Met Lys Arg Leu
385 390 395 400

Asn Leu Asp Arg Arg Glu Leu Ser Cys Leu Lys Ala Ile Ile Leu Tyr
405 410 415

Asn Pro Asp Ile Arg Gly Ile Lys Ser Arg Ala Glu Ile Glu Met Cys
420 425 430

Arg Glu Lys Val Tyr Ala Cys Leu Asp Glu His Cys Arg Leu Glu His
435 440 445

Pro Gly Asp Asp Gly Arg Phe Ala Gln Leu Leu Leu Arg Leu Pro Ala
450 455 460

Leu Arg Ser Ile Ser Leu Lys Cys Gln Asp His Leu Phe Leu Phe Arg
465 470 475 480

Ile Thr Ser Asp Arg Pro Leu Glu Glu Leu Phe Leu Glu Gln Leu Glu
485 490 495

Ala Pro Pro Pro Pro Gly Leu Ala Met Lys Leu Glu
500 505

<210> 49
<211> 555
<212> PRT
<213> Bombyx mori
<400> 49

Met His Glu Asp Ala Pro Lys Met Ser Ile Ala Gln Ser Leu Ala Ala
1 5 10 15

Ser Thr Ser Gln Pro Lys Gly Asp Ile Val Thr Glu Ile Pro Leu Glu
20 25 30

```
Phe Ala Met Ser Ser Met Glu Thr Lys Ser Ile Glu Thr Thr Asn Val
        35                  40                  45

Glu Leu Lys Ile Thr Tyr Val Asp Pro Thr Thr Gly Thr Gly Gly Glu
    50                  55                  60

Pro Gly Ala Tyr Leu Pro Thr Ala Gly Thr Val Cys Asp Gln Thr Asp
65                  70                  75                  80

Thr Lys Asp Val Ile Glu Glu Leu Cys Pro Val Cys Gly Asp Lys Val
                85                  90                  95

Ser Gly Tyr His Tyr Gly Leu Leu Thr Cys Glu Ser Cys Lys Gly Phe
            100                 105                 110

Phe Lys Arg Thr Val Gln Asn Lys Lys Val Tyr Thr Cys Val Ala Glu
        115                 120                 125

Arg Ala Cys His Ile Asp Lys Thr Gln Arg Lys Arg Cys Pro Phe Cys
    130                 135                 140

Arg Phe Gln Lys Cys Leu Asp Val Gly Met Lys Leu Glu Ala Val Arg
145                 150                 155                 160

Ala Asp Arg Met Arg Gly Gly Arg Asn Lys Phe Gly Pro Met Tyr Lys
                165                 170                 175

Arg Asp Arg Ala Arg Lys Leu Gln Met Met Arg Gln Arg Gln Ile Ala
            180                 185                 190

Val Gln Thr Leu Arg Gly Ser Leu Gly Asp Gly Gly Leu Val Leu Gly
        195                 200                 205

Phe Gly Ser Pro Tyr Thr Ala Val Ser Val Lys Gln Glu Ile Gln Ile
    210                 215                 220

Pro Gln Val Ser Ser Leu Thr Ser Ser Pro Glu Ser Ser Pro Gly Pro
225                 230                 235                 240

Ala Leu Leu Arg Ala Gln Pro Gln Pro Pro Gln Pro Pro Pro Pro Pro
                245                 250                 255

Thr His Asp Lys Trp Glu Ala His Ser Pro His Ser Ala Ser Pro Asp
            260                 265                 270

Ala Phe Thr Phe Asp Thr Gln Ser Asn Thr Ala Ala Thr Pro Ser Ser
        275                 280                 285

Thr Ala Glu Ala Thr Ser Thr Glu Thr Leu Arg Val Ser Pro Met Ile
    290                 295                 300

Arg Glu Phe Val Gln Thr Val Asp Asp Arg Glu Trp Gln Asn Ala Leu
305                 310                 315                 320

Phe Gly Leu Leu Gln Ser Gln Thr Tyr Asn Gln Cys Glu Val Asp Leu
                325                 330                 335

Phe Glu Leu Met Cys Lys Val Leu Asp Gln Asn Leu Phe Ser Gln Val
            340                 345                 350
```

Asp Trp Ala Arg Asn Thr Val Phe Phe Lys Tyr Leu Lys Val Asp Asp
355 360 365

Gln Met Lys Leu Leu Gln Asp Ser Trp Ser Val Met Leu Val Leu Asp
370 375 380

His Leu His Gln Arg Met His Asn Gly Leu Pro Asp Glu Thr Thr Leu
385 390 395 400

His Asn Gly Gln Lys Phe Asp Leu Leu Cys Leu Gly Leu Leu Gly Val
405 410 415

Pro Ser Leu Ala Asp His Phe Asn Glu Leu Gln Asn Lys Leu Ala Glu
420 425 430

Leu Lys Phe Asp Val Pro Asp Tyr Ile Cys Val Lys Phe Met Leu Leu
435 440 445

Leu Asn Pro Glu Val Arg Gly Ile Val Asn Val Lys Cys Val Arg Glu
450 455 460

Gly Tyr Gln Thr Val Gln Ala Ala Leu Leu Asp Tyr Pro Tyr Leu Leu
465 470 475 480

Ser Thr Ile Gln Asp Lys Phe Gly Lys Leu Val Met Val Val Pro Glu
485 490 495

Ile His Ala Leu Arg Leu Gly Glu Lys Ser Thr Cys Thr Ser Gly Ile
500 505 510

Val Gln Ala Arg His Leu Pro Arg Leu Phe Ser Trp Lys Cys Cys Thr
515 520 525

Gln Asn Ala Asn Leu Glu Val Pro Val Thr Asn Lys Val Glu Glu Leu
530 535 540

Arg Ser Ala Lys Pro Arg Arg His His Asn Lys
545 550 555

<210> 50
<211> 699
<212> PRT
<213> Manduca sexta
<400> 50

Met Thr Leu Val Met Ser Pro Asp Ser Ser Tyr Gly Arg Tyr Asp Ala
1 5 10 15

Pro Thr Pro Thr Asp Val Thr Ser Pro Val His Arg Glu Arg Glu Pro
20 25 30

Glu Leu His Ile Glu Phe Asp Gly Thr Thr Val Leu Cys Arg Val Cys
35 40 45

Gly Asp Lys Ala Ser Gly Phe His Tyr Gly Val His Ser Cys Glu Gly
50 55 60

Cys Lys Gly Phe Phe Arg Arg Ser Ile Gln Gln Lys Ile Gln Tyr Arg
65 70 75 80

```
Pro Cys Thr Lys Asn Gln Gln Cys Ser Ile Leu Arg Ile Asn Arg Asn
                85                  90                  95

Arg Cys Gln Tyr Cys Arg Leu Lys Lys Cys Ile Ala Val Gly Met Ser
            100                 105                 110

Arg Asp Ala Val Arg Phe Gly Arg Val Pro Lys Arg Glu Lys Ala Arg
        115                 120                 125

Ile Leu Ala Ala Met Gln Gln Ser Ser Thr Ser Arg Ala His Glu Gln
    130                 135                 140

Ala Ala Ala Ala Glu Leu Asp Asp Ala Pro Arg Leu Leu Ala Arg Val
145                 150                 155                 160

Val Arg Ala His Leu Asp Thr Cys Glu Phe Thr Arg Asp Arg Val Ala
                165                 170                 175

Ala Met Arg Ala Arg Ala Arg Asp Cys Pro Thr Tyr Ser Gln Pro Thr
            180                 185                 190

Leu Ala Cys Pro Leu Asn Pro Ala Pro Glu Leu Gln Ser Glu Lys Glu
        195                 200                 205

Phe Ser Gln Arg Phe Ala His Val Ile Arg Gly Val Ile Asp Phe Ala
    210                 215                 220

Gly Leu Ile Pro Gly Phe Gln Leu Leu Thr Gln Asp Asp Lys Phe Thr
225                 230                 235                 240

Leu Leu Lys Ser Gly Leu Phe Asp Ala Leu Phe Val Arg Leu Ile Cys
                245                 250                 255

Met Phe Asp Ala Pro Leu Asn Ser Ile Ile Cys Leu Asn Gly Gln Leu
            260                 265                 270

Met Lys Arg Asp Ser Ile Gln Ser Gly Ala Asn Ala Arg Phe Leu Val
        275                 280                 285

Asp Ser Thr Phe Lys Phe Ala Glu Arg Met Asn Ser Met Asn Leu Thr
    290                 295                 300

Asp Ala Glu Ile Gly Leu Phe Cys Ala Ile Val Leu Ile Thr Pro Asp
305                 310                 315                 320

Arg Pro Gly Leu Arg Asn Val Glu Leu Val Glu Arg Met His Thr Arg
                325                 330                 335

Leu Lys Ala Cys Leu Gln Thr Val Ile Ala Gln Asn Arg Pro Asp Arg
            340                 345                 350

Pro Gly Phe Leu Arg Glu Leu Met Asp Thr Leu Pro Asp Leu Arg Thr
        355                 360                 365

Leu Ser Thr Leu His Thr Glu Lys Leu Val Val Phe Arg Thr Glu His
    370                 375                 380

Lys Glu Leu Leu Arg Gln Gln Met Trp Ser Glu Glu Glu Ala Val Ser
385                 390                 395                 400
```

Trp Val Asp Ser Gly Ala Asp Glu Leu Ala Arg Ser Pro Ile Gly Ser
405 410 415

Val Ser Ser Ser Glu Ser Gly Glu Ala Val Gly Asp Cys Gly Thr Pro
420 425 430

Leu Leu Ala Ala Thr Leu Ala Gly Arg Arg Arg Leu Asp Ser Arg Gly
435 440 445

Ser Val Asp Glu Glu Ala Leu Gly Val Ala His Leu Ala His Asn Gly
450 455 460

Leu Thr Val Thr Pro Val Arg Gln Pro Pro Arg Tyr Arg Lys Leu Asp
465 470 475 480

Ser Pro Thr Asp Ser Gly Ile Glu Ser Gly Asn Glu Lys His Glu Arg
485 490 495

Ile Val Gly Thr Gly Ser Gly Cys Ser Ser Pro Arg Ser Ser Leu Glu
500 505 510

Glu His Asn Glu Asp Arg Arg Pro Pro Val Ser Ala Asp Asp Met Pro
515 520 525

Val Leu Lys Arg Val Leu Gln Ala Pro Pro Leu Tyr Gly Gly Thr Pro
530 535 540

Ser Leu Met Asp Glu Ala Tyr Arg Arg His Lys Lys Phe Arg Ala Leu
545 550 555 560

Arg Arg Asp Thr Gly Glu Ala Glu Ala Arg Thr Val Arg Pro Thr Pro
565 570 575

Ser Pro Gln Pro Gln His Pro His Pro Ala Asn Pro Ala His Pro Ala
580 585 590

His Ser Pro Arg Pro Gln Arg Ala Ser Leu Ser Ser Thr His Ser Val
595 600 605

Leu Ala Lys Ser Leu Met Glu Gly Pro Arg Met Thr Pro Glu Gln Leu
610 615 620

Lys Arg Thr Asp Ile Ile Gln Gln Tyr Met Arg Arg Gly Glu Ser Ser
625 630 635 640

Ala Pro Ala Glu Gly Cys Pro Leu Arg Ala Gly Gly Leu Leu Thr Cys
645 650 655

Tyr Arg Gly Ala Ser Pro Ala Pro Gln Pro Val Leu Ala Leu Gln Val
660 665 670

Asp Val Thr Asp Ala Pro Leu Asn Leu Ser Lys Lys Ser Pro Ser Pro
675 680 685

Pro Arg Thr Tyr Met Pro Gln Met Leu Glu Ala
690 695

<210> 51
<211> 1394
<212> PRT

<213> Drosophila melanogaster
<400> 51

Met Val Cys Ala Met Gln Glu Val Ala Ala Val Gln His Gln Gln Gln
1 5 10 15

Gln Gln Gln Leu Gln Leu Pro Gln Gln Gln Gln Gln Gln Gln Gln Thr
20 25 30

Thr Gln Gln Gln His Ala Thr Thr Ile Val Leu Leu Thr Gly Asn Gly
35 40 45

Gly Gly Asn Leu His Ile Val Ala Thr Pro Gln Gln His Gln Pro Met
50 55 60

His Gln Leu His His Gln His Gln His Gln His Gln His Gln Gln Gln
65 70 75 80

Ala Lys Ser Gln Gln Leu Lys Gln Gln His Ser Ala Leu Val Lys Leu
85 90 95

Leu Glu Ser Ala Pro Ile Lys Gln Gln Gln Gln Thr Pro Lys Gln Ile
100 105 110

Val Tyr Leu Gln Gln Gln Gln Gln Gln Pro Gln Arg Lys Arg Leu Lys
115 120 125

Asn Glu Ala Ala Ile Val Gln Gln Gln Gln Gln Thr Pro Ala Thr Leu
130 135 140

Val Lys Thr Thr Thr Thr Ser Asn Ser Asn Ser Asn Asn Thr Gln Thr
145 150 155 160

Thr Asn Ser Ile Ser Gln Gln Gln Gln Gln His Gln Ile Val Leu Gln
165 170 175

His Gln Gln Pro Ala Ala Ala Ala Thr Pro Lys Pro Cys Ala Asp Leu
180 185 190

Ser Ala Lys Asn Asp Ser Glu Ser Gly Ile Asp Glu Asp Cys Pro Asn
195 200 205

Ser Asp Glu Asp Cys Pro Asn Ala Asn Pro Ala Gly Thr Ser Leu Glu
210 215 220

Asp Ser Ser Tyr Glu Gln Tyr Gln Cys Pro Trp Lys Lys Ile Arg Tyr
225 230 235 240

Ala Arg Glu Leu Leu Lys Gln Arg Glu Leu Glu Gln Gln Gln Thr Thr
245 250 255

Gly Gly Ser Asn Ala Gln Gln Gln Val Glu Ala Lys Pro Ala Ala Ile
260 265 270

Pro Thr Ser Asn Ile Lys Gln Leu His Cys Asp Ser Pro Phe Ser Ala
275 280 285

Gln Thr His Lys Glu Ile Ala Asn Leu Leu Arg Gln Gln Ser Gln Gln
290 295 300

```
Gln Gln Val Val Ala Thr Gln Gln Gln Gln Gln Gln Gln Gln His
305                 310                 315                 320

Gln His Gln Gln Gln Arg Arg Asp Ser Ser Asp Ser Asn Cys Ser Leu
                325                 330                 335

Met Ser Asn Ser Ser Asn Ser Ser Ala Gly Asn Cys Cys Thr Cys Asn
            340                 345                 350

Ala Gly Asp Asp Gln Gln Leu Glu Glu Met Asp Glu Ala His Asp Ser
        355                 360                 365

Gly Cys Asp Asp Glu Leu Cys Glu Gln His His Gln Arg Leu Asp Ser
    370                 375                 380

Ser Gln Leu Asn Tyr Leu Cys Gln Lys Phe Asp Glu Lys Leu Asp Thr
385                 390                 395                 400

Ala Leu Ser Asn Ser Ser Ala Asn Thr Gly Arg Asn Thr Pro Ala Val
                405                 410                 415

Thr Ala Asn Glu Asp Ala Asp Gly Phe Phe Arg Arg Ser Ile Gln Gln
            420                 425                 430

Lys Ile Gln Tyr Arg Pro Cys Thr Lys Asn Gln Gln Cys Ser Ile Leu
        435                 440                 445

Arg Ile Asn Arg Asn Arg Cys Gln Tyr Cys Arg Leu Lys Lys Cys Ile
    450                 455                 460

Ala Val Gly Met Ser Arg Asp Ala Val Arg Phe Gly Arg Val Pro Lys
465                 470                 475                 480

Arg Glu Lys Ala Arg Ile Leu Ala Ala Met Gln Gln Ser Thr Gln Asn
                485                 490                 495

Arg Gly Gln Gln Arg Ala Leu Ala Thr Glu Leu Asp Asp Gln Pro Arg
            500                 505                 510

Leu Leu Ala Ala Val Leu Arg Ala His Leu Glu Thr Cys Glu Phe Thr
        515                 520                 525

Lys Glu Lys Val Ser Ala Met Arg Gln Arg Ala Arg Asp Cys Pro Ser
    530                 535                 540

Tyr Ser Met Pro Thr Leu Leu Ala Cys Pro Leu Asn Pro Ala Pro Glu
545                 550                 555                 560

Leu Gln Ser Glu Gln Glu Phe Ser Gln Arg Phe Ala His Val Ile Arg
                565                 570                 575

Gly Val Ile Asp Phe Ala Gly Met Ile Pro Gly Phe Gln Leu Leu Thr
            580                 585                 590

Gln Asp Asp Lys Phe Thr Leu Leu Lys Ala Gly Leu Phe Asp Ala Leu
        595                 600                 605

Phe Val Arg Leu Ile Cys Met Phe Asp Ser Ser Ile Asn Ser Ile Ile
    610                 615                 620
```

```
Cys Leu Asn Gly Gln Val Met Arg Arg Asp Ala Ile Gln Asn Gly Ala
625                 630                 635                 640

Asn Ala Arg Phe Leu Val Asp Ser Thr Phe Asn Phe Ala Glu Arg Met
                645                 650                 655

Asn Ser Met Asn Leu Thr Asp Ala Glu Ile Gly Leu Phe Cys Ala Ile
            660                 665                 670

Val Leu Ile Thr Pro Asp Arg Pro Gly Leu Arg Asn Leu Glu Leu Ile
        675                 680                 685

Glu Lys Met Tyr Ser Arg Leu Lys Gly Cys Leu Gln Tyr Ile Val Ala
    690                 695                 700

Gln Asn Arg Pro Asp Gln Pro Glu Phe Leu Ala Lys Leu Leu Glu Thr
705                 710                 715                 720

Met Pro Asp Leu Arg Thr Leu Ser Thr Leu His Thr Glu Lys Leu Val
                725                 730                 735

Val Phe Arg Thr Glu His Lys Glu Leu Leu Arg Gln Gln Met Trp Ser
            740                 745                 750

Met Glu Asp Gly Asn Asn Ser Asp Gly Gln Gln Asn Lys Ser Pro Ser
        755                 760                 765

Gly Ser Trp Ala Asp Ala Met Asp Val Glu Ala Ala Lys Ser Pro Leu
    770                 775                 780

Gly Ser Val Ser Ser Thr Glu Ser Ala Asp Leu Asp Tyr Gly Ser Pro
785                 790                 795                 800

Ser Ser Ser Gln Pro Gln Gly Val Ser Leu Pro Ser Pro Pro Gln Gln
                805                 810                 815

Gln Pro Ser Ala Leu Ala Ser Ser Ala Pro Leu Leu Ala Ala Thr Leu
            820                 825                 830

Ser Gly Gly Cys Pro Leu Arg Asn Arg Ala Asn Ser Gly Ser Ser Gly
        835                 840                 845

Asp Ser Gly Ala Ala Glu Met Asp Ile Val Gly Ser His Ala His Leu
    850                 855                 860

Thr Gln Asn Gly Leu Thr Ile Thr Pro Ile Val Arg His Gln Gln Gln
865                 870                 875                 880

Gln Gln Gln Gln Gln Gln Ile Gly Ile Leu Asn Asn Ala His Ser Arg
                885                 890                 895

Asn Leu Asn Gly Gly His Ala Met Cys Gln Gln Gln Gln Gln His Pro
            900                 905                 910

Gln Leu His His His Leu Thr Ala Gly Ala Ala Arg Tyr Arg Lys Leu
        915                 920                 925

Asp Ser Pro Thr Asp Ser Gly Ile Glu Ser Gly Asn Glu Lys Asn Glu
    930                 935                 940
```

Cys Lys Ala Val Ser Ser Gly Gly Ser Ser Ser Cys Ser Ser Pro Arg
945 950 955 960

Ser Ser Val Asp Asp Ala Leu Asp Cys Ser Asp Ala Ala Ala Asn His
965 970 975

Asn Gln Val Val Gln His Pro Gln Leu Ser Val Val Ser Val Ser Pro
980 985 990

Val Arg Ser Pro Gln Pro Ser Thr Ser Ser His Leu Lys Arg Gln Ile
995 1000 1005

Val Glu Asp Met Pro Val Leu Lys Arg Val Leu Gln Ala Pro Pro
1010 1015 1020

Leu Tyr Asp Thr Asn Ser Leu Met Asp Glu Ala Tyr Lys Pro His
1025 1030 1035

Lys Lys Phe Arg Ala Leu Arg His Arg Glu Phe Glu Thr Ala Glu
1040 1045 1050

Ala Asp Ala Ser Ser Ser Thr Ser Gly Ser Asn Ser Leu Ser Ala
1055 1060 1065

Gly Ser Pro Arg Gln Ser Pro Val Pro Asn Ser Val Ala Thr Pro
1070 1075 1080

Pro Pro Ser Ala Ala Ser Ala Ala Ala Gly Asn Pro Ala Gln Ser
1085 1090 1095

Gln Leu His Met His Leu Thr Arg Ser Ser Pro Lys Ala Ser Met
1100 1105 1110

Ala Ser Ser His Ser Val Leu Ala Lys Ser Leu Met Ala Glu Pro
1115 1120 1125

Arg Met Thr Pro Glu Gln Met Lys Arg Ser Asp Ile Ile Gln Asn
1130 1135 1140

Tyr Leu Lys Arg Glu Asn Ser Thr Ala Ala Ser Ser Thr Thr Asn
1145 1150 1155

Gly Val Gly Asn Arg Ser Pro Ser Ser Ser Ser Thr Pro Pro Pro
1160 1165 1170

Ser Ala Val Gln Asn Gln Gln Arg Trp Gly Ser Ser Ser Val Ile
1175 1180 1185

Thr Thr Thr Cys Gln Gln Arg Gln Gln Ser Val Ser Pro His Ser
1190 1195 1200

Asn Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
1205 1210 1215

Ser Ser Ser Ser Thr Ser Ser Asn Cys Ser Ser Ser Ser Ala Ser
1220 1225 1230

Ser Cys Gln Tyr Phe Gln Ser Pro His Ser Thr Ser Asn Gly Thr
1235 1240 1245

Ser Ala Pro Ala Ser Ser Ser Ser Gly Ser Asn Ser Ala Thr Pro
1250 1255 1260

Leu Leu Glu Leu Gln Val Asp Ile Ala Asp Ser Ala Gln Pro Leu
1265 1270 1275

Asn Leu Ser Lys Lys Ser Pro Thr Pro Pro Pro Ser Lys Leu His
1280 1285 1290

Ala Leu Val Ala Ala Ala Asn Ala Val Gln Arg Tyr Pro Thr Leu
1295 1300 1305

Ser Ala Asp Val Thr Val Thr Ala Ser Asn Gly Gly Ser Ser Val
1310 1315 1320

Gly Gly Gly Glu Ser Gly Arg Gln Gln Gln Ser Ala Gly Glu Cys
1325 1330 1335

Gly Leu Pro Gln Ser Gly Pro Glu Arg Arg Arg Ala Gln Gly Asn
1340 1345 1350

Ala Gly Gly Val Arg Ala Gly Gly Gly Arg Trp Phe Tyr Ala Glu
1355 1360 1365

Lys Trp Glu Arg Gln Arg Leu Gly Val Ala Val Gln Arg Ser Arg
1370 1375 1380

Lys Gln Asp His Leu Glu Arg Arg Glu Leu Asn
1385 1390

<210> 52
<211> 685
<212> PRT
<213> Manduca sexta
<400> 52

Met Val Arg Ala Met Ser Cys Gly Ala Glu Leu Arg Glu Arg His Ser
1 5 10 15

Val Leu Val Ser Met Leu Glu Ala Arg Arg Glu Ser Ser Asp Ser Gly
20 25 30

Cys Ser Ser Asp Asp Gly Ser Asp Val Glu Arg Asp Cys Lys Cys Arg
35 40 45

Cys Asp Pro Gln Gly Phe Phe Arg Arg Ser Ile Gln Gln Lys Ile Gln
50 55 60

Tyr Arg Pro Cys Thr Lys Asn Gln Gln Cys Ser Ile Leu Arg Ile Asn
65 70 75 80

Arg Asn Arg Cys Gln Tyr Cys Arg Leu Lys Lys Cys Ile Ala Val Gly
85 90 95

Met Ser Arg Asp Ala Val Arg Phe Gly Arg Val Pro Lys Arg Glu Lys
100 105 110

Ala Arg Ile Leu Ala Ala Met Gln Gln Ser Ser Thr Ser Arg Ala His
115 120 125

```
Glu Gln Ala Ala Ala Ala Glu Leu Asp Asp Ala Pro Arg Leu Leu Ala
    130                 135                 140

Arg Val Val Arg Ala His Leu Asp Thr Cys Glu Phe Thr Arg Asp Arg
145                 150                 155                 160

Val Ala Ala Met Arg Ala Arg Ala Arg Asp Cys Pro Thr Tyr Ser Gln
                165                 170                 175

Pro Thr Leu Ala Cys Pro Leu Asn Pro Ala Pro Glu Leu Gln Ser Glu
            180                 185                 190

Lys Glu Phe Ser Gln Arg Phe Ala His Val Ile Arg Gly Val Ile Asp
        195                 200                 205

Phe Ala Gly Leu Ile Pro Gly Phe Gln Leu Leu Thr Gln Asp Asp Lys
    210                 215                 220

Phe Thr Leu Leu Lys Ser Gly Leu Phe Asp Ala Leu Phe Val Arg Leu
225                 230                 235                 240

Ile Cys Met Phe Asp Ala Pro Leu Asn Ser Ile Ile Cys Leu Asn Gly
                245                 250                 255

Gln Leu Met Lys Arg Asp Ser Ile Gln Ser Gly Ala Asn Ala Arg Phe
            260                 265                 270

Leu Val Asp Ser Thr Phe Lys Phe Ala Glu Arg Met Asn Ser Met Asn
        275                 280                 285

Leu Thr Asp Ala Glu Ile Gly Leu Phe Cys Ala Ile Val Leu Ile Thr
    290                 295                 300

Pro Asp Arg Pro Gly Leu Arg Asn Val Glu Leu Val Glu Arg Met His
305                 310                 315                 320

Thr Arg Leu Lys Ala Cys Leu Gln Thr Val Ile Ala Gln Asn Arg Pro
                325                 330                 335

Asp Arg Pro Gly Phe Leu Arg Glu Leu Met Asp Thr Leu Pro Asp Leu
            340                 345                 350

Arg Thr Leu Ser Thr Leu His Thr Glu Lys Leu Val Val Phe Arg Thr
        355                 360                 365

Glu His Lys Glu Leu Leu Arg Gln Gln Met Trp Ser Glu Glu Glu Ala
    370                 375                 380

Val Ser Trp Val Asp Ser Gly Ala Asp Glu Leu Ala Arg Ser Pro Ile
385                 390                 395                 400

Gly Ser Val Ser Ser Ser Glu Ser Gly Glu Ala Val Gly Asp Cys Gly
                405                 410                 415

Thr Pro Leu Leu Ala Ala Thr Leu Ala Gly Arg Arg Arg Leu Asp Ser
            420                 425                 430

Arg Gly Ser Val Asp Glu Glu Ala Leu Gly Val Ala His Leu Ala His
        435                 440                 445
```

Asn Gly Leu Thr Val Thr Pro Val Arg Gln Pro Pro Arg Tyr Arg Lys
450 455 460

Leu Asp Ser Pro Thr Asp Ser Gly Ile Glu Ser Gly Asn Glu Lys His
465 470 475 480

Glu Arg Ile Val Gly Thr Gly Ser Gly Cys Ser Ser Pro Arg Ser Ser
485 490 495

Leu Glu Glu His Asn Glu Asp Arg Arg Pro Pro Val Ser Ala Asp Asp
500 505 510

Met Pro Val Leu Lys Arg Val Leu Gln Ala Pro Pro Leu Tyr Gly Gly
515 520 525

Thr Pro Ser Leu Met Asp Glu Ala Tyr Arg Arg His Lys Lys Phe Arg
530 535 540

Ala Leu Arg Arg Asp Thr Gly Glu Ala Glu Ala Arg Thr Val Arg Pro
545 550 555 560

Thr Pro Ser Pro Gln Pro Gln His Pro His Pro Ala Asn Pro Ala His
565 570 575

Pro Ala His Ser Pro Arg Pro Gln Arg Ala Ser Leu Ser Ser Thr His
580 585 590

Ser Val Leu Ala Lys Ser Leu Met Glu Gly Pro Arg Met Thr Pro Glu
595 600 605

Gln Leu Lys Arg Thr Asp Ile Ile Gln Gln Tyr Met Arg Arg Gly Glu
610 615 620

Ser Ser Ala Pro Ala Glu Gly Cys Pro Leu Arg Ala Gly Gly Leu Leu
625 630 635 640

Thr Cys Tyr Arg Gly Ala Ser Pro Ala Pro Gln Pro Val Leu Ala Leu
645 650 655

Gln Val Asp Val Thr Asp Ala Pro Leu Asn Leu Ser Lys Lys Ser Pro
660 665 670

Ser Pro Pro Arg Thr Tyr Met Pro Gln Met Leu Glu Ala
675 680 685

<210> 53
<211> 1443
<212> PRT
<213> Drosophila melanogaster
<400> 53

Met His Gly Gly Gly Pro Gly Ser Ser Gly Ser Asn Ile Ile Arg Arg
1 5 10 15

Ser Ser Gly Ser Phe Pro Gly Ser Gly Ser Gly Ser Ala Ser Lys Leu
20 25 30

Ile Lys Thr Glu Pro Ile Asp Phe Glu Met Leu His Leu Glu Glu Asn
35 40 45

```
Glu Arg Gln Gln Asp Ile Glu Arg Glu Pro Ser Ser Ser Asn Ser Asn
    50                  55                  60

Ser Asn Ser Asn Ser Leu Thr Pro Gln Arg Tyr Thr His Val Gln Val
65                  70                  75                  80

Gln Thr Val Pro Pro Arg Gln Pro Thr Gly Leu Thr Thr Pro Gly Gly
                85                  90                  95

Thr Gln Lys Val Ile Leu Thr Pro Arg Val Glu Tyr Val Gln Gln Arg
            100                 105                 110

Ala Thr Ser Ser Thr Gly Gly Gly Met Lys His Val Tyr Ser Gln Gln
        115                 120                 125

Gln Gly Thr Ala Ala Ser Arg Ser Ala Pro Pro Glu Thr Thr Ala Leu
    130                 135                 140

Leu Thr Thr Thr Ser Gly Thr Pro Gln Ile Ile Ile Thr Arg Thr Leu
145                 150                 155                 160

Pro Ser Asn Gln His Leu Ser Arg Arg His Ser Ala Ser Pro Ser Ala
                165                 170                 175

Leu His His Tyr Gln Gln Gln Gln Gln Pro Gln Arg Gln Gln Ser Pro
            180                 185                 190

Pro Pro Leu His His Gln Gln Gln Gln Gln Gln Gln His Val Arg Val
        195                 200                 205

Ile Arg Asp Gly Arg Leu Tyr Asp Glu Ala Thr Val Val Val Ala Ala
    210                 215                 220

Arg Arg His Ser Val Ser Pro Pro Pro Leu His His His Ser Arg Ser
225                 230                 235                 240

Ala Pro Val Ser Pro Val Ile Ala Arg Arg Gly Gly Ala Ala Ala Tyr
                245                 250                 255

Met Asp Gln Gln Tyr Gln Gln Arg Gln Thr Pro Pro Leu Ala Pro Pro
            260                 265                 270

Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Pro Gln Gln Gln
        275                 280                 285

Gln Gln Gln Tyr Ile Ser Thr Gly Val Pro Pro Pro Thr Ala Ala Ala
    290                 295                 300

Arg Lys Phe Val Val Ser Thr Ser Thr Arg His Val Asn Val Ile Ala
305                 310                 315                 320

Ser Asn His Phe Gln Gln Gln Gln Gln Gln His Gln Ala Gln Gln His
                325                 330                 335

Gln Gln Gln His Gln Gln His Val Ile Ala Ser Val Ser Ser Ser Ser
            340                 345                 350

Ser Ser Ser Ala Ile Gly Ser Gly Gly Ser Ser Ser Ser His Ile Phe
        355                 360                 365
```

```
Arg Thr Pro Val Val Ser Ser Ser Ser Ser Asn Met His His Gln
    370                 375                 380

Gln Gln Gln Gln Gln Gln Gln Ser Ser Leu Gly Asn Ser Val Met Arg
385                 390                 395                 400

Pro Pro Pro Pro Pro Pro Pro Pro Lys Val Lys His Ala Ser Ser Ser
                405                 410                 415

Ser Ser Gly Asn Ser Ser Ser Ser Asn Thr Asn Asn Ser Ser Ser Ser
            420                 425                 430

Ser Asn Gly Glu Glu Pro Ser Ser Ser Ile Pro Asp Leu Glu Phe Asp
        435                 440                 445

Gly Thr Thr Val Leu Cys Arg Val Cys Gly Asp Lys Ala Ser Gly Phe
    450                 455                 460

His Tyr Gly Val His Ser Cys Glu Gly Cys Lys Gly Phe Phe Arg Arg
465                 470                 475                 480

Ser Ile Gln Gln Lys Ile Gln Tyr Arg Pro Cys Thr Lys Asn Gln Gln
                485                 490                 495

Cys Ser Ile Leu Arg Ile Asn Arg Asn Arg Cys Gln Tyr Cys Arg Leu
            500                 505                 510

Lys Lys Cys Ile Ala Val Gly Met Ser Arg Asp Ala Val Arg Phe Gly
        515                 520                 525

Arg Val Pro Lys Arg Glu Lys Ala Arg Ile Trp Arg Pro Cys Asn Arg
    530                 535                 540

Ala Pro Arg Ile Ala Ala Ser Ser Asp Pro Ser Pro Pro Ser Trp Met
545                 550                 555                 560

Thr Ser His Ala Ser Ser Pro Pro Cys Cys Cys Ala His Leu Glu Thr
                565                 570                 575

Cys Glu Phe Thr Lys Glu Lys Val Ser Ala Met Arg His Gly Arg Gly
            580                 585                 590

Leu Pro Ser Thr Pro Cys His Thr Ser Gly Leu Ser Ala Glu Pro Ala
        595                 600                 605

Pro Glu Leu Gln Ser Glu Gln Glu Phe Ser Gln Arg Phe Ala His Val
    610                 615                 620

Ile Arg Gly Val Ile Asp Phe Ala Gly Met Ile Pro Gly Phe Gln Leu
625                 630                 635                 640

Leu Thr Gln Asp Asp Lys Phe Thr Leu Leu Lys Ala Gly Leu Phe Asp
                645                 650                 655

Ala Leu Phe Val Arg Leu Ile Cys Met Phe Asp Ser Ser Ile Asn Ser
            660                 665                 670

Ile Ile Cys Leu Asn Gly Gln Val Met Arg Arg Asp Ala Ile Gln Asn
        675                 680                 685
```

Gly Ala Asn Ala Arg Phe Leu Val Asp Ser Thr Phe Asn Phe Ala Glu
690 695 700

Arg Met Asn Ser Met Asn Leu Thr Asp Ala Glu Ile Gly Leu Phe Cys
705 710 715 720

Ala Ile Val Leu Ile Thr Pro Asp Arg Pro Gly Leu Arg Asn Leu Glu
725 730 735

Leu Ile Glu Lys Met Tyr Ser Arg Leu Lys Gly Cys Leu Gln Tyr Ile
740 745 750

Val Ala Gln Asn Arg Pro Asp Gln Pro Glu Phe Leu Ala Lys Leu Leu
755 760 765

Glu Thr Met Pro Asp Leu Arg Thr Leu Ser Thr Leu His Thr Glu Lys
770 775 780

Leu Val Val Phe Arg Thr Glu His Lys Glu Leu Leu Arg Gln Gln Met
785 790 795 800

Trp Ser Met Glu Asp Gly Asn Asn Ser Asp Gly Gln Gln Asn Lys Ser
805 810 815

Pro Ser Gly Ser Trp Ala Asp Ala Met Asp Val Glu Ala Ala Lys Ser
820 825 830

Pro Leu Gly Ser Val Ser Ser Thr Glu Ser Ala Asp Leu Asp Tyr Gly
835 840 845

Ser Pro Ser Ser Ser Gln Pro Gln Gly Val Ser Leu Pro Ser Pro Pro
850 855 860

Gln Gln Gln Pro Ser Ala Leu Ala Ser Ser Ala Pro Leu Leu Ala Ala
865 870 875 880

Thr Leu Ser Gly Gly Cys Pro Leu Arg Asn Arg Ala Asn Ser Gly Ser
885 890 895

Ser Gly Asp Ser Gly Ala Ala Glu Met Asp Ile Val Gly Ser His Ala
900 905 910

His Leu Thr Gln Asn Gly Leu Thr Ile Thr Pro Ile Val Arg His Gln
915 920 925

Gln Gln Gln Gln Gln Gln Gln Gln Ile Gly Ile Leu Asn Asn Ala His
930 935 940

Ser Arg Asn Leu Asn Gly Gly His Ala Met Cys Gln Gln Gln Gln Gln
945 950 955 960

His Pro Gln Leu His His His Leu Thr Ala Gly Ala Ala Arg Tyr Arg
965 970 975

Lys Leu Asp Ser Pro Thr Asp Ser Gly Ile Glu Ser Gly Asn Glu Lys
980 985 990

Asn Glu Cys Lys Ala Val Ser Ser Gly Gly Ser Ser Ser Cys Ser Ser
995 1000 1005

Pro Arg Ser Ser Val Asp Asp Ala Leu Asp Cys Ser Asp Ala Ala
1010 1015 1020

Ala Asn His Asn Gln Val Val Gln His Pro Gln Leu Ser Val Val
1025 1030 1035

Ser Val Ser Pro Val Arg Ser Pro Gln Pro Ser Thr Ser Ser His
1040 1045 1050

Leu Lys Arg Gln Ile Val Glu Asp Met Pro Val Leu Lys Arg Val
1055 1060 1065

Leu Gln Ala Pro Pro Leu Tyr Asp Thr Asn Ser Leu Met Asp Glu
1070 1075 1080

Ala Tyr Lys Pro His Lys Lys Phe Arg Ala Leu Arg His Arg Glu
1085 1090 1095

Phe Glu Thr Ala Glu Ala Asp Ala Ser Ser Ser Thr Ser Gly Ser
1100 1105 1110

Asn Ser Leu Ser Ala Gly Ser Pro Arg Gln Ser Pro Val Pro Asn
1115 1120 1125

Ser Val Ala Thr Pro Pro Pro Val Ala Ala Ser Ala Ala Ala Gly
1130 1135 1140

Asn Pro Ala Gln Ser Gln Leu His Met His Leu Thr Arg Ser Ser
1145 1150 1155

Pro Lys Ala Ser Met Ala Ser Ser His Ser Val Leu Ala Lys Ser
1160 1165 1170

Leu Met Ala Glu Pro Arg Met Thr Pro Glu Gln Met Lys Arg Ser
1175 1180 1185

Asp Ile Ile Gln Asn Tyr Leu Lys Arg Glu Asn Ser Thr Ala Ala
1190 1195 1200

Ser Ser Thr Thr Asn Gly Leu Gly Asn Arg Ser Pro Ser Ser Ser
1205 1210 1215

Ser Thr Pro Pro Pro Ser Val Gln Asn Gln Gln Arg Trp Gly Ser
1220 1225 1230

Ser Ser Val Ile Thr Thr Thr Cys Gln Gln Arg Gln Gln Ser Val
1235 1240 1245

Ser Pro His Ser Asn Gly Ser Ser Ser Ser Ser Ser Ser Ser Ser
1250 1255 1260

Ser Ser Ser Ser Ser Ser Ser Ser Thr Ser Ser Asn Cys Ser Ser
1265 1270 1275

Ser Ser Ala Ser Ser Cys Gln Tyr Phe Gln Ser Pro His Ser Thr
1280 1285 1290

Ser Ile Gly Thr Gly Glu Pro Asp Gly Ala Pro Val Arg Asp Arg
1295 1300 1305

Thr Ala Pro Arg Pro Cys Trp Asn Cys Arg Trp Thr Leu Leu Thr
1310 1315 1320

Arg Arg Thr Ser Gln Phe Val Gln Glu Ile Ala His Ala Ala Ala
1325 1330 1335

Gln Gln Ala Ala Arg Ser Gly Gly Arg Arg Gln Cys Arg Ser Lys
1340 1345 1350

Val Ser His Ile Val Arg Arg Arg His Ser Asp Ser Leu Gln Trp
1355 1360 1365

Arg Ser Ser Val Gly Gly Gly Glu Ser Gly Ala Gln Gln Gln Ser
1370 1375 1380

Ala Gly Glu Cys Gly Leu Pro Gln Ser Gly Pro Glu Arg Arg Arg
1385 1390 1395

Ala Gln Gly Asn Ala Gly Gly Val Arg Ala Gly Gly Gly Arg Trp
1400 1405 1410

Phe Tyr Ala Glu Lys Trp Glu Arg Gln Arg Leu Gly Val Ala Val
1415 1420 1425

Gln Arg Ser Arg Lys Gln Asp His Leu Glu Arg Arg Glu Leu Asn
1430 1435 1440

<210> 54
<211> 690
<212> PRT
<213> Choristoneura fumiferana
<400> 54

Met Thr Leu Val Met Ser Pro Asp Ser Ser Tyr Gly Arg Tyr Asp Ala
1 5 10 15

Gln Pro Pro Val Asp Gly Gly Met Val Asn Pro Val His Arg Glu Arg
20 25 30

Glu Pro Glu Leu His Ile Glu Phe Asp Gly Thr Thr Val Leu Cys Arg
35 40 45

Val Cys Gly Asp Lys Ala Ser Gly Phe His Tyr Gly Val His Ser Cys
50 55 60

Glu Gly Cys Lys Gly Phe Phe Arg Arg Ser Ile Gln Gln Lys Ile Gln
65 70 75 80

Tyr Arg Pro Cys Thr Lys Asn Gln Gln Cys Ser Ile Leu Arg Ile Asn
85 90 95

Arg Asn Arg Cys Gln Tyr Cys Arg Leu Lys Lys Cys Ile Ala Val Gly
100 105 110

Met Ser Arg Asp Ala Val Arg Phe Gly Arg Val Pro Lys Arg Glu Lys
115 120 125

Ala Arg Ile Leu Ala Ala Met Gln Gln Ser Ser Ser Ser Arg Ala His
130 135 140

```
Glu Gln Ala Ala Ala Ala Glu Leu Asp Asp Ala Pro Arg Leu Leu Ala
145                 150                 155                 160

Arg Val Val Arg Ala His Leu Asp Thr Cys Glu Phe Thr Arg Asp Arg
                165                 170                 175

Val Ala Ala Met Arg Ala Arg Ala Arg Asp Cys Pro Thr Tyr Ser Gln
            180                 185                 190

Pro Thr Leu Ala Cys Pro Leu Asn Pro Ala Pro Glu Leu Gln Ser Glu
        195                 200                 205

Lys Glu Phe Ser Gln Arg Phe Ala His Val Ile Arg Gly Val Ile Asp
    210                 215                 220

Phe Ala Gly Leu Ile Pro Gly Phe Gln Leu Leu Thr Gln Asp Asp Lys
225                 230                 235                 240

Phe Thr Leu Leu Lys Ser Gly Leu Phe Asp Ala Leu Phe Val Arg Leu
                245                 250                 255

Ile Cys Met Phe Asp Ala Pro Leu Asn Ser Ile Ile Cys Leu Asn Gly
            260                 265                 270

Gln Leu Met Lys Arg Asp Ser Ile Gln Ser Gly Ala Asn Ala Arg Phe
        275                 280                 285

Leu Val Asp Ser Thr Phe Lys Phe Ala Glu Arg Met Asn Ser Met Asn
    290                 295                 300

Leu Thr Asp Ala Glu Ile Gly Leu Phe Cys Ala Ile Val Leu Ile Thr
305                 310                 315                 320

Pro Asp Arg Pro Gly Leu Arg Asn Ile Glu Leu Val Glu Arg Met His
                325                 330                 335

Ala Arg Leu Lys Ser Cys Leu Gln Thr Val Ile Ala Gln Asn Arg Ala
            340                 345                 350

Asp Arg Pro Gly Phe Leu Arg Glu Leu Met Asp Thr Leu Pro Asp Leu
        355                 360                 365

Arg Thr Leu Ser Thr Leu His Thr Glu Lys Leu Val Val Phe Arg Thr
    370                 375                 380

Glu His Lys Glu Leu Leu Arg Gln Gln Met Trp Gly Asp Glu Glu Val
385                 390                 395                 400

Cys Pro Trp Ala Asp Ser Gly Val Asp Asp Ser Ala Arg Ser Pro Leu
                405                 410                 415

Gly Ser Val Ser Ser Ser Glu Ser Gly Glu Ala Pro Ser Asp Cys Gly
            420                 425                 430

Thr Pro Leu Leu Ala Ala Thr Leu Ala Gly Arg Arg Arg Leu Asp Ser
        435                 440                 445

Arg Gly Ser Val Asp Glu Glu Ala Leu Gly Val Ala His Leu Ala His
    450                 455                 460
```

Asn Gly Leu Thr Val Thr Pro Val Arg Pro Pro Pro Arg Tyr Arg Lys
465 470 475 480

Leu Asp Ser Pro Thr Asp Ser Gly Ile Glu Ser Gly Asn Glu Lys His
485 490 495

Glu Arg Ile Val Gly Pro Gly Ser Gly Cys Ser Ser Pro Arg Ser Ser
500 505 510

Leu Glu Glu His Met Glu Asp Arg Arg Pro Leu Ala Ala Asp Asp Met
515 520 525

Pro Val Leu Lys Arg Val Leu Gln Ala Pro Pro Leu Tyr Asp Ala Ser
530 535 540

Ser Leu Met Asp Glu Ala Tyr Lys Pro His Lys Lys Phe Arg Ala Met
545 550 555 560

Arg Arg Asp Thr Gly Glu Ala Glu Ala Arg Pro Met Arg Pro Thr Pro
565 570 575

Ser Pro Gln Pro Met His Pro His Pro Gly Ser Pro Ala His Pro Ala
580 585 590

His Pro Ala His Ser Pro Arg Pro Leu Arg Ala Pro Leu Ser Ser Thr
595 600 605

His Ser Val Leu Ala Lys Ser Leu Met Glu Gly Pro Arg Met Thr Pro
610 615 620

Glu Gln Leu Lys Arg Thr Asp Ile Ile Gln Gln Tyr Met Arg Arg Gly
625 630 635 640

Glu Ala Gly Glu Glu Cys Arg Ala Gly Leu Leu Leu Tyr Arg Gly Ala
645 650 655

Ser Pro Leu Gln Val Asp Val Ala Asp Ala Pro Gln Pro Leu Asn Leu
660 665 670

Ser Lys Lys Ser Pro Ser Pro Pro Arg Ser Phe Met Pro Pro Met Leu
675 680 685

Glu Ala
690

<210> 55
<211> 711
<212> PRT
<213> Galleria mellonella
<400> 55

Met Thr Leu Val Met Ser Pro Asp Ser Ser Tyr Gly Arg Tyr Asp Ala
1 5 10 15

Pro Ala Pro Ala Asp Asn Arg Ile Met Ser Pro Val His Lys Glu Arg
20 25 30

Glu Pro Glu Leu His Ile Glu Phe Asp Gly Thr Thr Val Leu Cys Arg
35 40 45

```
Val Cys Gly Asp Lys Ala Ser Gly Phe His Tyr Gly Val His Ser Cys
    50                  55                  60

Glu Gly Cys Lys Gly Phe Phe Arg Arg Ser Ile Gln Gln Lys Ile Gln
65                  70                  75                  80

Tyr Arg Pro Cys Thr Lys Asn Gln Gln Cys Ser Ile Leu Arg Ile Asn
                85                  90                  95

Arg Asn Arg Cys Gln Tyr Cys Arg Leu Lys Lys Cys Ile Ala Val Gly
            100                 105                 110

Met Ser Arg Asp Ala Val Arg Phe Gly Arg Val Pro Lys Arg Glu Lys
        115                 120                 125

Ala Arg Ile Leu Ala Ala Met Gln Ser Ser Thr Thr Arg Ala His Glu
    130                 135                 140

Gln Ala Ala Ala Ala Glu Leu Asp Asp Gly Pro Arg Leu Leu Ala Arg
145                 150                 155                 160

Val Val Arg Ala His Leu Asp Thr Cys Glu Phe Thr Arg Asp Arg Val
                165                 170                 175

Ala Ala Met Arg Asn Gly Ala Arg Asp Cys Pro Thr Tyr Ser Gln Pro
            180                 185                 190

Thr Leu Ala Cys Pro Leu Asn Pro Ala Pro Glu Leu Gln Ser Glu Lys
        195                 200                 205

Glu Phe Ser Gln Arg Phe Ala His Val Ile Arg Gly Val Ile Asp Phe
    210                 215                 220

Ala Gly Leu Ile Pro Gly Phe Gln Leu Leu Thr Gln Asp Asp Lys Phe
225                 230                 235                 240

Thr Leu Leu Lys Ser Gly Leu Phe Asp Ala Leu Phe Val Arg Leu Ile
                245                 250                 255

Cys Met Phe Asp Ala Pro Leu Asn Ser Ile Ile Cys Leu Asn Gly Gln
            260                 265                 270

Leu Met Lys Arg Asp Ser Ile Gln Ser Gly Ala Asn Ala Arg Phe Leu
        275                 280                 285

Val Asp Ser Thr Phe Lys Phe Ala Glu Arg Met Asn Ser Met Asn Leu
    290                 295                 300

Thr Asp Ala Glu Ile Gly Leu Phe Cys Ala Ile Val Leu Ile Thr Pro
305                 310                 315                 320

Asp Arg Pro Gly Leu Arg Asn Val Glu Leu Val Glu Arg Met His Ser
                325                 330                 335

Arg Leu Lys Ser Cys Leu Gln Thr Val Ile Ala Gln Asn Arg Ser Asp
            340                 345                 350

Gly Pro Gly Phe Leu Arg Glu Leu Met Asp Thr Leu Pro Asp Leu Arg
        355                 360                 365
```

```
Thr Leu Ser Thr Leu His Thr Glu Lys Leu Val Val Phe Arg Thr Glu
    370                 375                 380

His Lys Glu Leu Leu Arg Gln Gln Met Trp Val Glu Asp Glu Gly Ala
385                 390                 395                 400

Leu Trp Ala Asp Ser Gly Ala Asp Asp Ser Ala Arg Ser Pro Ile Gly
                405                 410                 415

Ser Val Ser Ser Ser Glu Ser Ser Glu Thr Thr Gly Asp Cys Gly Thr
            420                 425                 430

Pro Leu Leu Ala Ala Thr Leu Ala Gly Arg Arg Arg Leu Asp Ser Arg
        435                 440                 445

Gly Ser Val Asp Glu Glu Ala Leu Gly Val Ala His Leu Ala His Asn
    450                 455                 460

Gly Leu Thr Val Thr Pro Val Arg Pro Pro Pro Arg Tyr Arg Lys Leu
465                 470                 475                 480

Asp Ser Pro Thr Asp Ser Gly Ile Glu Ser Gly Asn Glu Lys His Glu
                485                 490                 495

Arg Ile Val Gly Pro Glu Ser Gly Cys Ser Ser Pro Arg Ser Ser Leu
            500                 505                 510

Glu Glu His Ser Asp Asp Arg Arg Pro Ile Ala Pro Ala Asp Asp Met
        515                 520                 525

Pro Val Leu Lys Arg Val Leu Gln Ala Pro Pro Leu Tyr Asp Ala Ser
    530                 535                 540

Ser Leu Met Asp Glu Ala Tyr Lys Pro His Lys Lys Phe Arg Ala Met
545                 550                 555                 560

Arg Arg Asp Thr Trp Ser Glu Ala Glu Ala Arg Pro Gly Arg Pro Thr
                565                 570                 575

Pro Ser Pro Gln Pro Pro His His Pro His Pro Ala Ser Pro Ala His
            580                 585                 590

Pro Ala His Ser Pro Arg Pro Ile Arg Ala Pro Leu Ser Ser Thr His
        595                 600                 605

Ser Val Leu Ala Lys Ser Leu Met Glu Gly Pro Arg Met Thr Pro Glu
    610                 615                 620

Gln Leu Lys Arg Thr Asp Ile Ile Gln Gln Tyr Met Arg Arg Gly Glu
625                 630                 635                 640

Thr Gly Ala Pro Thr Glu Gly Cys Pro Leu Arg Ala Gly Gly Leu Leu
                645                 650                 655

Thr Cys Phe Arg Gly Ala Ser Pro Ala Pro Gln Pro Val Ile Ala Leu
            660                 665                 670

Gln Val Asp Val Ala Glu Thr Asp Ala Pro Gln Pro Leu Asn Leu Ser
        675                 680                 685
```

Lys Lys Ser Pro Ser Pro Ser Pro Pro Pro Pro Pro Pro Arg Ser Tyr
690 695 700

Met Pro Pro Met Leu Pro Ala
705 710

<210> 56
<211> 711
<212> PRT
<213> Metapenaeus ensis
<400> 56

Met Thr Leu Val Met Ser Pro Asp Ser Ser Tyr Gly Arg Tyr Asp Ala
1 5 10 15

Pro Ala Pro Ala Asp Asn Arg Ile Met Ser Pro Val His Lys Glu Arg
20 25 30

Glu Pro Glu Leu His Ile Glu Phe Asp Gly Thr Thr Val Leu Cys Arg
35 40 45

Val Cys Gly Asp Lys Ala Ser Gly Phe His Tyr Gly Val His Ser Cys
50 55 60

Glu Gly Cys Lys Gly Phe Phe Arg Arg Ser Ile Gln Gln Lys Ile Gln
65 70 75 80

Tyr Arg Pro Cys Thr Lys Asn Gln Gln Cys Ser Ile Leu Arg Ile Asn
85 90 95

Arg Asn Arg Cys Gln Tyr Cys Arg Leu Lys Lys Cys Ile Ala Val Gly
100 105 110

Met Ser Arg Asp Ala Val Arg Phe Gly Arg Val Pro Lys Arg Glu Lys
115 120 125

Ala Arg Ile Leu Ala Ala Met Gln Ser Ser Thr Thr Arg Ala His Glu
130 135 140

Gln Ala Ala Ala Ala Glu Leu Asp Asp Gly Pro Arg Leu Leu Ala Arg
145 150 155 160

Val Val Arg Ala His Leu Asp Thr Cys Glu Phe Thr Arg Asp Arg Val
165 170 175

Ala Ala Met Arg Asn Gly Ala Arg Asp Cys Pro Thr Tyr Ser Gln Pro
180 185 190

Thr Leu Ala Cys Pro Leu Asn Pro Ala Pro Glu Leu Gln Ser Glu Lys
195 200 205

Glu Phe Ser Gln Arg Phe Ala His Val Ile Arg Gly Val Ile Asp Phe
210 215 220

Ala Gly Leu Ile Pro Gly Phe Gln Leu Leu Thr Gln Asp Asp Lys Phe
225 230 235 240

Thr Leu Leu Lys Ser Gly Leu Phe Asp Ala Leu Phe Val Arg Leu Ile
245 250 255

Cys Met Phe Asp Ala Pro Leu Asn Ser Ile Ile Cys Leu Asn Gly Gln
260 265 270

Leu Met Lys Arg Asp Ser Ile Gln Ser Gly Ala Asn Ala Arg Phe Leu
275 280 285

Val Asp Ser Thr Phe Lys Phe Ala Glu Arg Met Asn Ser Met Asn Leu
290 295 300

Thr Asp Ala Glu Ile Gly Leu Phe Cys Ala Ile Val Leu Ile Thr Pro
305 310 315 320

Asp Arg Pro Gly Leu Arg Asn Val Glu Leu Val Glu Arg Met His Ser
325 330 335

Arg Leu Lys Ser Cys Leu Gln Thr Val Ile Ala Gln Asn Arg Ser Asp
340 345 350

Gly Pro Gly Phe Leu Arg Glu Leu Met Asp Thr Leu Pro Asp Leu Arg
355 360 365

Thr Leu Ser Thr Leu His Thr Glu Lys Leu Val Val Phe Arg Thr Glu
370 375 380

His Lys Glu Leu Leu Arg Gln Gln Met Trp Val Glu Asp Glu Gly Ala
385 390 395 400

Leu Trp Ala Asp Ser Gly Ala Asp Asp Ser Ala Arg Ser Pro Ile Gly
405 410 415

Ser Val Ser Ser Ser Glu Ser Ser Glu Thr Thr Gly Asp Cys Gly Thr
420 425 430

Pro Leu Leu Ala Ala Thr Leu Ala Gly Arg Arg Arg Leu Asp Ser Arg
435 440 445

Gly Ser Val Asp Glu Glu Ala Leu Gly Val Ala His Leu Ala His Asn
450 455 460

Gly Leu Thr Val Thr Pro Val Arg Pro Pro Pro Arg Tyr Arg Lys Leu
465 470 475 480

Asp Ser Pro Thr Asp Ser Gly Ile Glu Ser Gly Asn Glu Lys His Glu
485 490 495

Arg Ile Val Gly Pro Glu Ser Gly Cys Ser Ser Pro Arg Ser Ser Leu
500 505 510

Glu Glu His Ser Asp Asp Arg Arg Pro Ile Ala Pro Ala Asp Asp Met
515 520 525

Pro Val Leu Lys Arg Val Leu Gln Ala Pro Pro Leu Tyr Asp Ala Ser
530 535 540

Ser Leu Met Asp Glu Ala Tyr Lys Pro His Lys Lys Phe Arg Ala Met
545 550 555 560

Arg Arg Asp Thr Trp Ser Glu Ala Glu Ala Arg Pro Gly Arg Pro Thr
565 570 575

Pro Ser Pro Gln Pro Pro His His Pro His Pro Ala Ser Pro Ala His
580 585 590

Pro Ala His Ser Pro Arg Pro Ile Arg Ala Pro Leu Ser Ser Thr His
595 600 605

Ser Val Leu Ala Lys Ser Leu Met Glu Gly Pro Arg Met Thr Pro Glu
610 615 620

Gln Leu Lys Arg Thr Asp Ile Ile Gln Gln Tyr Met Arg Arg Gly Glu
625 630 635 640

Thr Gly Ala Pro Thr Glu Gly Cys Pro Leu Arg Ala Gly Gly Leu Leu
645 650 655

Thr Cys Phe Arg Gly Ala Ser Pro Ala Pro Gln Pro Val Ile Ala Leu
660 665 670

Gln Val Asp Val Ala Glu Thr Asp Ala Pro Gln Pro Leu Asn Leu Ser
675 680 685

Lys Lys Ser Pro Ser Pro Ser Pro Pro Pro Pro Pro Pro Arg Ser Tyr
690 695 700

Met Pro Pro Met Leu Pro Ala
705 710